META-ANALYSIS OF HYDRAULIC FRACTURE CONDUCTIVITY DATA

By

Mohammed Rashnur Rahman

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of

Master of Science

in

Petroleum Engineering

University of Alaska Fairbanks

December 2017

APPROVED:

Obadare O. Awoleke, Committee Chair
Scott Goddard, Committee Co-Chair
Mohabbat Ahmadi, Committee Member
Abhijit Dandekar, Chair
Department of Petroleum Engineering
Douglas J. Goering, Dean
College of Engineering and Mines
Michael Castellini, *Dean of the Graduate School*

ABSTRACT

Previous empirical models of propped fracture conductivity are based either on data sourced from single investigations or on data not in the public domain. In this work, statistically rigorous models of propped fracture conductivity are developed using a database of fracture conductivity experiments reported in technical literature over the last 40 years. The database contains the results from about 2700 experimental runs. Propped fracture conductivity is the dependent variable and proppant types, mesh size, proppant concentration, formation hardness, closure stress, formation temperature, and polymer concentration are the independent variables. The mother database is partitioned into subsets; that is different databases with each daughter database having complete information in relation to the dependent and independent variables. As a result, the number of independent variables included in the daughter databases varied from three to six. Seventy percent of the data was used to develop the models while 30% of the data was used to validate them. First, fixed effect models were developed using regression analysis. Afterwards, three, four and five factor models were compared for two types of proppant: sand and ceramic proppant. The five factor model appeared to be the most prominent one.

The analysis was further carried out using five factors of these two types of proppant. Mixed effect modeling was employed because of the disparate sources of the data and also the temporal diversity of the dataset. The mixed effect model appeared to be the better than the fixed effect model while compared the error terms. Also, because the mother database contained some missing values, two statistical imputation approaches were employed to predict the missing values which are categorical imputation and multiple imputation using chained equations. Imputations are employed because it is speculated that a model developed using a large number of data points should provide better predictions. Generally, the mean squared error (MSE) is less in the mixed effect model for sand and in the categorical imputation model for ceramic proppant. But, to be more precise on the performance of the models, model predictions were compared with an existing propped fracture conductivity model and different case histories published in literature. Subsequently, the models of this research can be arranged in order of predictive performance: multiple imputation model, mixed effect model, fixed effect/categorical imputation model. The results also indicate that mesh size, closure stress, formation hardness, and proppant concentration significantly affect fracture conductivity from a statistical point of view. Formation temperature and polymer concentration affect conductivity negatively but they were not statistically significant.

Engineers will have access to a propped fracture conductivity database based on experiments reported over the past 40 years in technical literature. Engineers can use the models developed based on this database to generate statistical distributions of propped fracture conductivity for a variety of proppant characteristics and formation conditions. The models presented here are based on data from experimental investigations in different laboratories thereby reducing the bias that may be present in single laboratory investigations.

DEDICATION

First of all, I dedicate this work to Almighty Allah. I would also like to dedicate this work to my late father who throughout his life encouraged and supported me and my pursuits in higher education. Last but not least, I would like to dedicate this work to my mother and my family.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

NOMENCLATURE

BHN = Formation Harndess (Brinnel Hardness Number, kg/mm^2)

h = Height of the Core

b = Width of Fracture Face

L = Fracture Length

L = Length of the Core

M = Molecular Weight

q = Flow rate

q = Flow Rate through the Core

R = Universal Gas Constant

T = Temperature (^{0}F)

w = Fracture Width

Z = Compressibility Factor

ΔP = Pressure Differential Fracture Conductivity through the Core

C_p = Proppant Concentration (lb/ft^2)

C_{pl} = Polymer Concentration (lb/gal)

$Fc,\ k_f w$ = Fracture Conductivity (md-ft)

P_{ms} = Proppant Mesh Size (inch)

p_1 = Inlet Pressure

p_2 = Outlet Pressure

t_P = Types of Proppant

σ_c = Closure Stress

β = Non-Darcy Flow Coefficient

μ = Fluid Viscosity

ρ = Fluid Density

ACKNOWLEDGEMENTS

First of all, I would like to thank Dr. Obadare Awoleke for giving me the opportunity to pursue my MS in petroleum engineering under his close supervision. His direction and support throughout my MS really helped me to complete my study. He has also aided me with his intellectual thoughts and guidance for the completion of my thesis. His hard work to facilitate a graduate student's thesis is really commendable.

Next, I would like to thank Dr. Scott Goddard for sharing his insights on statistics with me. As a student of petroleum engineering, it has been really hard for me to use statistical tools in this particular field. His advice on statistics undoubtedly assisted me to make conclusive statements in my thesis.

I would also like to thank Dr. Mohabbat Ahmadi for being a member of my research committee. His support and encouragement also helped to develop the confidence necessary to complete my thesis.

Next, I would like to convey my gratitude to the Department of Petroleum Engineering, University of Alaska Fairbanks, and the chair Dr. Abhijit Dandekar for selecting me as a graduate student. I would like to thank Kushagra Saxena and Aditya Nikam for helping me with data collection at the initial stage. I thank Neeraj Mahajan for his tremendous support of sending materials from distance. I also thank Keith Robertson for providing me with important references. I would also like to thank my family and friends for their comforting words throughout this process.

Last but not least, I would like to thank Dr. Trina Mamoon for being a guardian, mentor, and friend during my journey at Fairbanks.

Chapter 1 INTRODUCTION

1.1 Background

Hydraulic fracturing (fracking) is a stimulation technique in reservoirs and is used to enhance well productivity as well as ultimate hydrocarbon recovery. The process includes pumping highly viscous fluids which contains mainly water, proppants, and other chemical additives at high pressure into the reservoir. The fluid is pumped at pressures exceeding the formation fracturing pressure to break down rock in order to create artificial fractures. When the hydraulic pressure is released, the proppants keep the fractures open and new flow channels are created for reservoir fluid to flow easier. Fracture conductivity is the capacity of the fracture to transmit fluid from the reservoir to the wellbore. It is defined as the formation permeability multiplied by width of the fracture. There are several important factors that can affect fracture conductivity. These factors include the types of proppant used in the slurry, proppant particle size, closure stress, proppant concentration, reservoir temperature, polymer loading, gas rate, the presence of breaker, time, formation hardness, and proppant embedment.

Hydraulic fracturing was initially attempted in the 1940s as field experiments. Commercial application of this technique by several service companies started in the 1950's. In recent years, a large number of hydraulic fracturing treatments have been carried out all over the world most expansively in the USA. The U.S. Energy Information Administration conducted a study in 2016 on oil and gas production in the United States from 2000-2015. They concluded that in 2015, approximately 51% of oil and 67% of natural gas production came from hydraulically fractured wells. **Figure 1-1** and **Figure 1-2** clearly show the rapid increase in oil and gas production caused by the use of hydraulic fracturing techniques in the USA. This large number of hydraulic fracturing treatments in the USA has increased the competition among the both operator companies and service providers. Due to this large number of hydraulic fracturing treatment, it is a priority to reduce the cost. Therefore, it is needed to accurately predict hydraulic fracture conductivity using correlations that are applicable for wide ranges of affecting factors. The primary objective has therefore become to optimize the treatment by developing empirical models that can be used to determine the effect of various factors on fracture conductivity. Several fracture conductivity models have been developed by different researchers. Most of them were developed either using a limited number of experiments (Awoleke et al., 2012) or considering only a limited number of factors (Barree et al., 2016). Therefore, it has become important to develop models of fracture

conductivity considering wide ranges of data and also a good number of dominant affecting factors. The major objective of this research is to provide more general fracture conductivity models based on large amounts of experimental data that are collected from disparate sources. This research will provide different rigorous fracture conductivity models using advanced statistical tools. This study also includes statistical methods for the engineers to develop empirical models, deal with missing data and asses uncertainty.

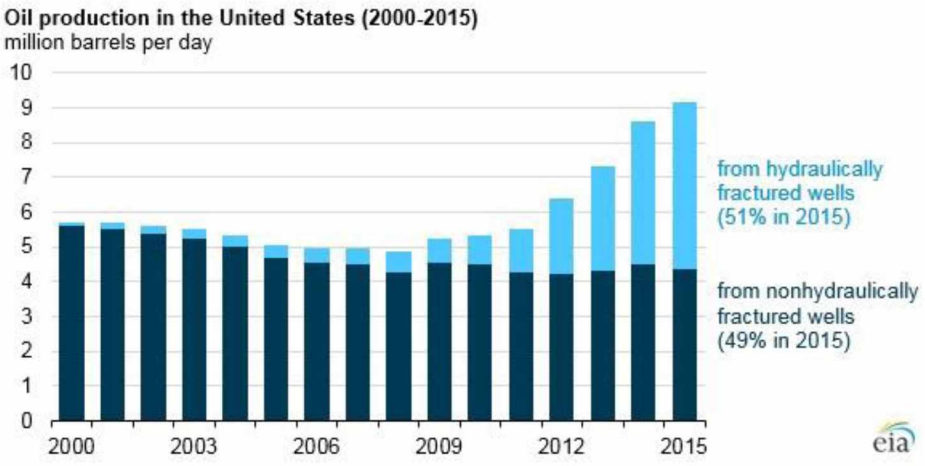

Figure 1-1: Oil Production in the USA for 2000-2015 (Source: US Energy Information Administration, 2016)

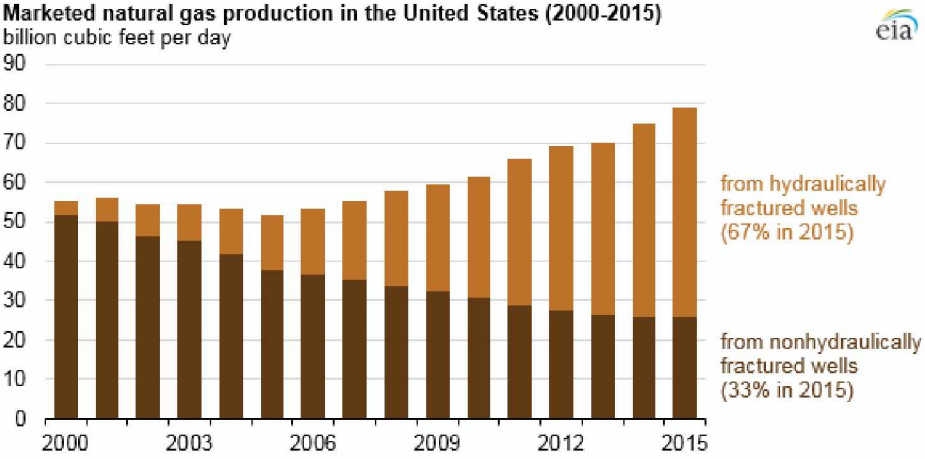

Figure 1-2: Natural Gas Production in the USA for 2000-2015 (Source: US Energy Information Administration, 2016)

1.2 Problem Description

Over the last 45 years, researchers have developed several fracture conductivity models which are used in reservoir models to predict well performance. Although the documented effects of the controlling factors are found to be similar in most research, the magnitudes describing the effect of each factor controlling fracture conductivity varies depending on the source of data. Additionally, most of the research was performed concentrating on a small number of factors affecting fracture conductivity (Barree et al., 2016). Therefore, more general conductivity models are required for future prediction. It is also important to identify the most dominant affecting factors on conductivity and include them in the models.

This study represents different empirical models of hydraulic fracture conductivity from a wide range of experimental data. However, gathering a wide range of data by conducting experiments in laboratory is both time consuming and expensive. Therefore, these experimental data are collated from fifteen research publications and one dissertation on hydraulic fracturing. So in this study, a large amount of data is being used to develop conductivity models. Most of the dominant factors, which have been the concerns for researchers (Anderson et al., 1989), are included in this study as well. This study represents a methodology of the application of statistical analysis in the field of petroleum engineering. Statistically robust models are developed here by checking model adequacy and estimating uncertainty. The method takes into account the fact that the data are collated from disparate sources. This study also presents statistical methods for engineers to deal with missing data and use them to develop empirical correlations. This study emphasizes detailed fracture conductivity modeling of two types of proppant that are sand and ceramic proppants. But the correlations are applicable for a wide range of sand or ceramic materials that have been used so far for laboratory measurements of fracture conductivity and there is no need to specify a particular sand or ceramic proppant to use the correlations for prediction.

1.3 Research Objectives

The empirical models developed in this study are based on advanced statistical tools. This research has the following objectives:

- Build a mother database of the dependent and independent variables. From the literature, we are considering proppant mesh size, closure stress, proppant concentration, reservoir temperature, polymer loading, formation hardness, and types of proppant as independent variables and fracture conductivity as the dependent variable.
- Divide the mother database into subsets that are different daughter databases with each daughter database having complete information in relation to the dependent and independent variables because the mother database will necessarily contain missing information.
- Develop fracture conductivity models based on each daughter database by employing regression analysis.
- Employ a mixed effects model to develop other models of fracture conductivity in order to consider the disparate source of these experimental data.
- Employ different forms of statistical imputation to develop models to predict fracture conductivity.
- Compare all the different models with previously established models and also compare with case histories.
- Provide recommendation for future work.

Chapter 2 LITERATURE REVIEW

2.1 Fracture Conductivity Calculation and Measurement

Fracture conductivity is defined as the fracture permeability multiplied by the width of the fracture. Usually, it is measured in millidarcy-feet (md-ft). Engineers use it as a measure of the capacity of the fracture to transmit fluid. Laboratory measurement of proppant pack conductivity/fracture conductivity is based on Darcy's law and the basic fracture conductivity equation (Anderson et al., 1989) is given below:

$$kw_f = \frac{q\mu L}{h\Delta P} \tag{2.1}$$

In the literature, there are some other modified equations that are used to calculate fracture conductivity. Hill et al. (2007) modified Forchheimer's equation for fluid flow through a porous medium and derived the following:

$$\frac{(p_1{}^2 - p_2{}^2)Mb}{2ZRTL\mu\rho q} = \frac{1}{k_f w} + \frac{\beta\rho q}{w^2\mu h} \tag{2.2}$$

Duenckel et al. (2016) provided a brief history of the development of proppant conductivity measurement in the laboratory. The very first conductivity cell was the Cooke conductivity cell (1973). The Cooke conductivity cell is shown in **Figure 2-1**. Fracture conductivity experiments up to a closure stress limit of 5000 psi and a temperature limit of 300 degF were carried out using the Cooke's cell. To avoid non-Darcy effect, permeability was measured at low flow rates. Brine was injected into the cell at elevated temperatures in order to simulate reservoir conditions. Later Gulf and Mobil developed a linear flow cell known as the linear flow conductivity test unit (Gidley et al., 1989), as shown in **Figure 2-2**. This conductivity test unit could successfully measure the fracture width for any linear configuration.

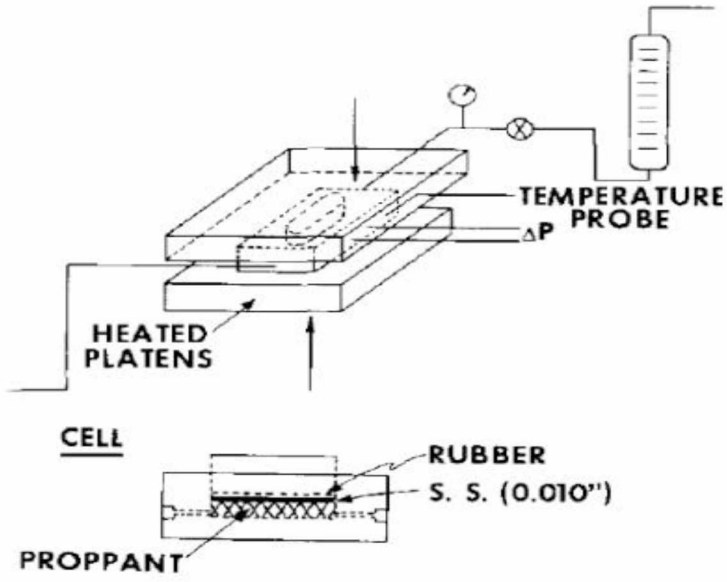

Figure 2-1: Cooke Conductivity Cell (Cooke, 1973)

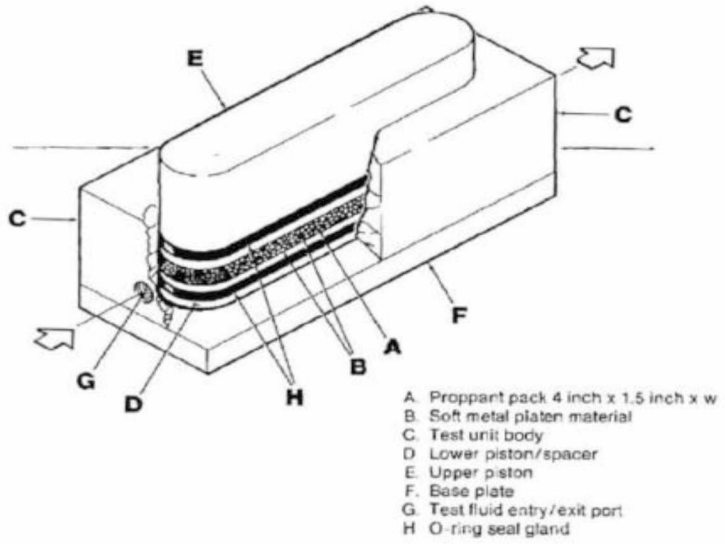

Figure 2-2: Linear flow conductivity test unit (Gidley et al., 1989)

API issued the first industry standard conductivity cell to measure proppant pack conductivity in October of 1989. It was known as the API fracture conductivity unit (API, 1989) shown in **Figure 2-3**. It is used mainly to measure short-term conductivity. It consists of a metal platen, lower and upper piston, and a porous metal filter. Experiments are conducted at following conditions:

- A 10-inch square flow path (1.5 inch width and 7.0 inch length cell).
- Deionized or distilled water.
- Experiments conducted at ambient temperature: 75^0F +/- 5^0F.
- Proppant confined between steel platens with a concentration of 2 lbm/ft^2.
- Proppant stressed for 15 minute periods at each stress.
- Recommended stresses for varying size and type of proppant.
- Stresses ranged from 1 to 14K psi.

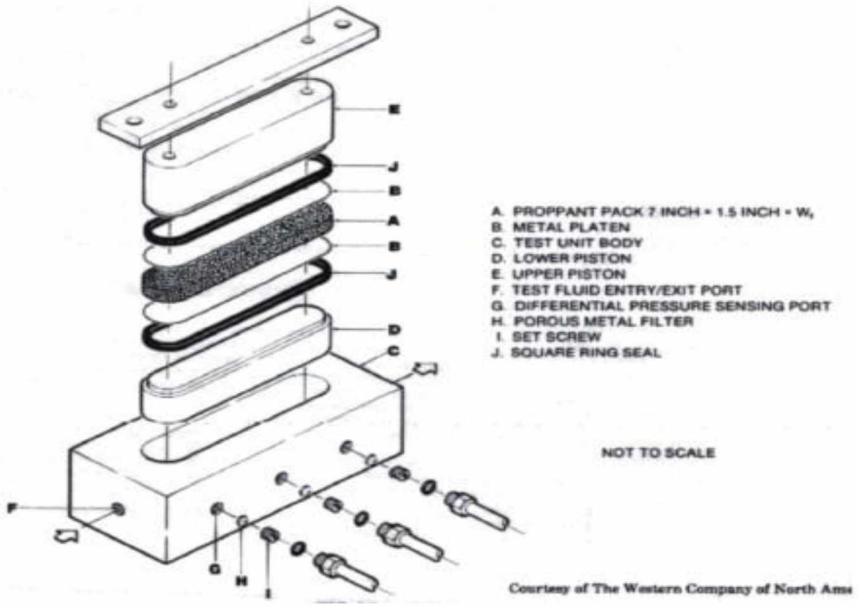

Figure 2-3: API conductivity unit (API, 1989)

In this conductivity cell, proppant is confined between steel platens. As a result, there is no chance of proppant embedment on the surface of the platen. This does not adequately mimic reservoir conditions. In reality, we know proppant embedment occurs. So, conductivity can be over-estimated by the experiment. A comparison was made with steel platen and Ohio sandstone at different stresses which showed 46% reduction in conductivity at 10000 psi stress level (Penny, 1987) as shown in **Figure 2-4**. To provide an accurate estimation of conductivity considering proppant embedment, steel platen was replaced by sandstone. The following changes are also made in the API conductivity cell:

- Replacement of deionized water by 2% KCL.

- Long-term conductivity measurement.
- Test temperatures from 150^0F to 250^0F.

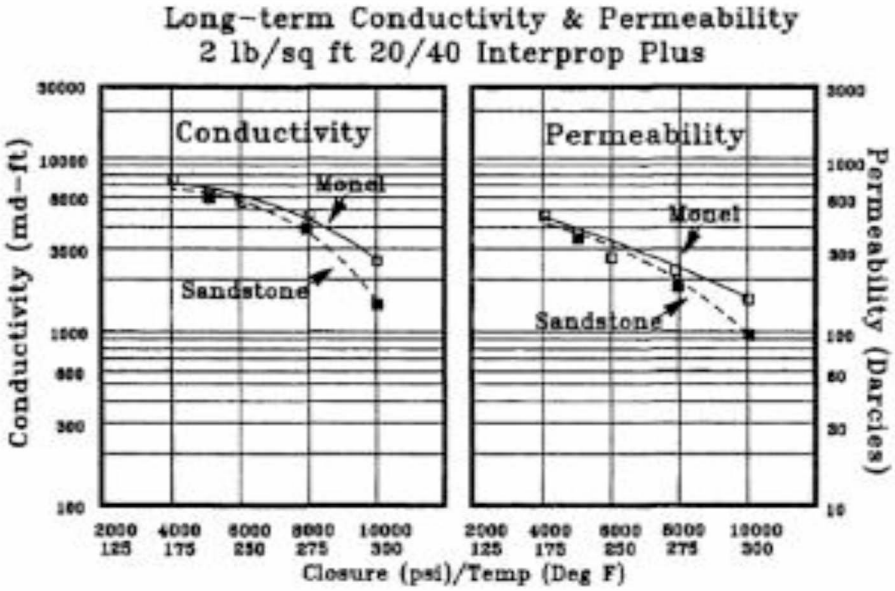

Figure 2-4: Conductivity of proppant confined in between steel platens and sandstone (Penny, 1987)

More modifications were made to the standard API conductivity cell. These modifications are stated briefly below:

- The use of advanced transducers in order to be able to accurately measure pressure difference.
- The use of a pluviator replaced human cell loading, delivered better crush results, and uniform distribution across conductivity cell.
- Deeper Groove reduced uncertainties in width measurements, stress uniformly applied to the proppant pack.
- Still shim reduced sandstone core failure.
- Modified cell length (9.5 inches) saved the cell from deformation.

In recent days, Marpaung et al. (2008) conducted experiments using two side pistons with Vilton polypack seals to keep the rock sample in place. They cut the rock samples in a rectangular shape with dimensions of 7 inch long, 1.65 inch wide and 3 inch thick. The 3 inch thickness provided

more controlled leakoff and allowed more filtration around in the rock around the fracture. Kakkar (2016) used a Hassler type core holder to perform fracture conductivity experiments and applied the confining pressure radially on the core holder. Stainless steel accumulators were used to contain the test fluid.

2.2 Factors Affecting Fracture Conductivity

In this section, we will discuss the various factors that can affect fracture conductivity development.

2.2.1 Proppant Particle Size

Using an optimal proppant particle size is very important in designing a fracturing treatment. The proppant particle size must be small enough to enter the fracture. A minimum fracture width should be in the range of two to three times the largest diameter of a proppant particle (Anderson et al., 1989). It was also mentioned that to use 8/16, 20-40 and 40-70 mesh proppants, a minimum fracture width of 0.187, 0.066 and 0.033 in is required respectively. Fracture conductivity increases with increase in proppant particle size, although it has been established that smaller proppants can be transported to the depths of a fracture more easily than larger proppants. On the other hand, larger proppants can form a partial monolayer which provides significant conductivity at lower stress levels (Kamenov, 2013). At a higher stress, however, proppant particles are crushed and proppant strength decreases as the size increases. Fracture conductivity also depends on proppant size distribution, proppant shape, and surface areas.

2.2.2 Closure Stress

The effect of closure stress is well defined by the researchers from different laboratory experiments. Usually, closure stress is determined by subtracting flowing bottomhole pressure from minimum horizontal stress. This closure stress on proppants causes crushing of the proppants, which results in reducing the particle size and increases the surface area. Both of these effectively reduce the permeability of a propped fracture. The compressive strength of a proppant must be higher than the closure stress in order to resist the crushing. Closure stress also causes compaction of particle bed which reduces the porosity. It is also a cause of proppant embedment into the formation which in turn causes a reduction in fracture width and subsequently fracture permeability.

2.2.3 Proppant Concentration

Proppant concentration is a measure of the amount of proppant per unit area of one side of a fracture wall. It is usually expressed in pounds of proppant per square foot of a fracture wall. However, proppant concentration can also be considered at the pump which is expressed by pounds per gallon. Hydraulic fracture conductivity generally increases with increasing proppant concentration if we do not consider the partial mono-layer concept. The increase in proppant concentration increases fracture flow capacity, minimizes crushing of the sand, and increases proppant system tolerances to fines (Coulter and Wells, 1972).

2.2.4 Reservoir Temperature

Reservoir temperature is another important factor that should be considered for a hydraulic fracturing treatment design. McDaniel (1987) evidently stated that a hydraulic fracture conductivity design requires conductivity data that are a function of reservoir temperature. At high temperature, the proppant itself can be altered as a result of thermal degradation. The less durable proppants face significant loss in strength at high temperature and afterwards, the fracture conductivity decreases. Temperature also affects the break time of fracturing fluid and the amount of residue left in the fracture. John et al. (2017) found that the HPG polymer degradation rate increases with increase in temperature.

2.2.5 Polymer Loading

The primary reason for using polymer in fracturing fluid is to increase the viscosity of the fluid. Many researchers have conducted studies to evaluate effects of polymer loading on fracture conductivity over the last three decades. Cooke (1975) indicated that the residue from the polymer used in the experiments causes a reduction in fracture conductivity.

2.2.6 Formation Hardness

Formation hardness is another important factor which has adverse effects on fracture conductivity. Vlis et al. (1975) unsuccessfully attempted to establish theoretical relationships between fracture conductivity and proppant type, size, and concentration as well as closure stress and formation hardness. Later, an extensive experimental program resulted in a set of graphs using Brinell hardness number (BHN) as a simple parameter of rock-strength classification. These graphs were thereafter used to establish empirical relationships for fracture conductivity. The use of these empirical correlations is limited to sand, walnut hulls and glass beads of few specific mesh sizes.

Therefore, more general correlations are needed for extensive use. BHN is a parameter of formation hardness quality expressed in kilograms per square of a millimeter.

2.2.7 Types of Proppant

Usually, there are three main types of proppant available for use in the industry: sand, resin-coated sand, and ceramic proppant. The hierarchy of conductivity according to the proppant types is shown in **Figure 2-5**. From **Figure 2-5** it is clear that using ceramic proppant should provide the maximum conductivity whereas sand will provide the minimum conductivity.

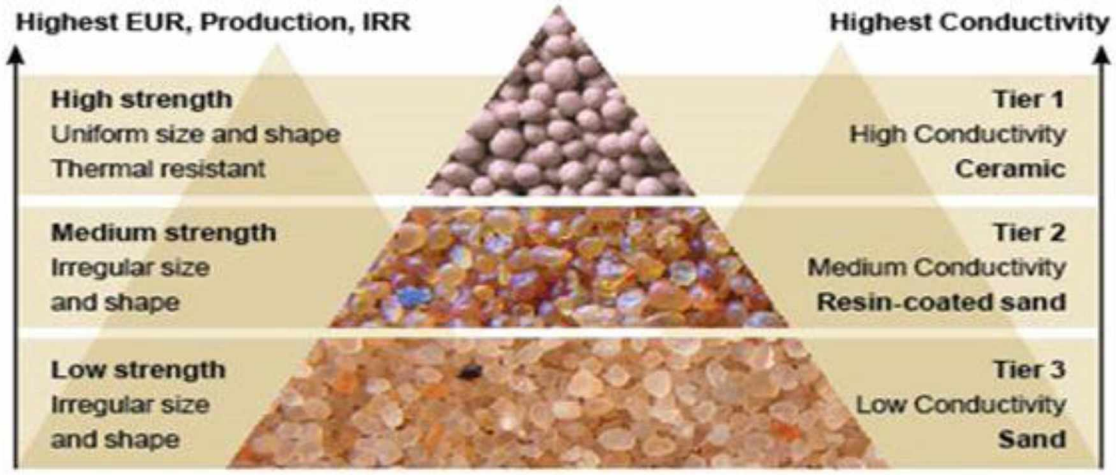

Figure 2-5: Hierarchy of Conductivity (Source: GY Ceramic Proppant Co. Ltd., 2012)

In early days, hydraulic fracturing treatments were done using sand such as Jordan Ottawa Sand, Alteveer Sand, Hickory Sand, and Brown Sand, all of which are known as the lowest strength proppant. Several attempts were made to improve the quality of sand as most of these failed to achieve necessary conductivity at higher stress levels. Later on, resin-coated sand and ceramic proppants were used in order to get high performance from hydraulically fractured wells. Resin-coated sand can be found in two forms: curable and procured and it is usually employed for intermediate closure stress applications. Based on density, three types of ceramic proppants are available for use: namely, lightweight ceramics, intermediate-density ceramics, and high-density ceramics. Ceramic proppants are known as high strength proppants. This type of proppant is manufactured to withstand the greatest closure stress. Sintered Bauxite, Zirconia, Mullite are example ceramic proppants. To optimize the cost, selection of a proppant type for a hydraulic

fracturing treatment is very important. Using high strength proppants might not be feasible all the time depending on the formation closure stress. At lower closure stress, using sand as a proppant type can be both productive and economical.

2.2.8 Other Factors

There are certain other factors which can affect fracture conductivity both positively and negatively. Researchers are still conducting laboratory experiments to find the full effect of these factors, which means there is still a lack of experimental data available in the literature to evaluate these factors.

2.2.8.1 Proppant Embedment

After the proppants are pumped into the fractures, proppants might penetrate the walls of the fracture which leads to a reduction of fracture width. As a result fracture conductivity is decreased. Li et al. (2015) developed mathematical models to calculate proppant embedment and fracture conductivity. They concluded that proppant embedment increases with increases in closure stress, the diameter of proppants, and elastic modulus of proppant. On the other hand, it decreases with increasing elastic modulus of coalbed.

2.2.8.2 Fracturing Fluid Residue

Sometimes deposition of fracturing fluid residue in the fractures causes reduction of pore spaces. It can cause a reduction in fracture conductivity. Using high polymer concentration in the fracturing fluid increases the amount of the residues. Cooke (1975) stated that the amount of residue in the fluid and fraction of residue retained in the fracture as fluid leaks off are the factors determining the severity of the reduction in fracture conductivity.

2.2.8.3 Gas Rate and Presence of Breaker

Awoleke et al. (2016) found that an increase in gas rate can decrease fracture conductivity. On the other hand, adding a breaker leads to an increase in conductivity. In addition, they mentioned that the physical mechanism behind the effect of gas rate is still unclear and the effect of the breaker is less noticeable at high temperatures because of polymer dehydration.

2.3 Previous Fracture Conductivity Experiments and Models

A large number of laboratory conductivity tests have been conducted over the last 45 years to understand the effects of these factors. Cooke (1973) conducted laboratory tests to understand the effect of temperature and closure stress. The apparatus was used to measure the permeability of

vertical fractures packed with several layers of brittle proppants at different stress levels. Two types of fracturing fluid were used in his experiments: water based and oil based. He stated that permeability of some fractures is negatively affected by stress in the presence of high-temperature brine. At 250^0F there was a significant difference in permeability and stress behavior in the presence of water and oil.

Cooke (1975) developed theoretical models to evaluate the effect of fracturing fluid on conductivity. The major finding was the adverse effect of gel residue on conductivity. Polymer concentration, breaker type, and concentration are the important factors controlling the amount of residue. Vlis et al. (1975) conducted experiments using sands, glass beads, and walnut hulls as proppants at different concentration and stress levels and developed a set of empirical relations determining fracture conductivity by using BHN as a rock strength parameter. Volk et al. (1981) developed empirical equations to assess proppant embedment on sand and shale by using sintered bauxite as proppant. Cutler et al. (1985) measured fracture conductivity at different stress levels by using wide varieties of resin-coated and ceramic proppants. At higher stress levels, Bauxite and Zirconia A provided the highest conductivity out of all the proppants used in the experiments. Cutler et al. (1985) recommended these two proppants for deep wells and resin-coated proppant for higher proppant flow back. It has been demonstrated by Parker and McDaniel (1987) that filter cakes formed by the fracturing fluid on fractures can have harmful effects on conductivity while 20/40 Ottawa sand was used as proppant. They reported that using an intermediate strength proppant is required even at low closure stress. And that regardless of proppant types, it is necessary to pump proppants at high concentration to get the desired conductivity in the fractures.

McDaniel (1986) studied proppant conductivity using five proppants of different composition. The main variables of the study were: comparison between short-term and long-term conductivity under different test conditions, the effects of the change from ambient to elevated temperatures on fracture conductivity, and effects of brine flow at high temperature and stress on fracture conductivity. He found a significant reduction in long-term conductivity at elevated temperature and stress. Penny (1987) demonstrated effects of closure stress, temperature, proppant embedment, fracturing fluid interactions, and fluid loss additives on short-term as well as long-term conductivity. The importance of considering time as a factor to evaluate effects of stress on conductivity was reported by Much and Penny (1987).

Fredd et al. (2001) conducted a series of laboratory experiments using cores obtained from Texas Cotton Valley formation to investigate effects of certain properties on conductivity. Experiments were conducted for both smooth and rough fracture surfaces. They stated that displacement of rough fracture walls can provide increased conductivity in the absence of proppant. While applying water fracturing treatments, the success will depend significantly on formation properties such as the degree of fracture displacement, the size and distribution of asperities, and rock mechanical properties. They also recommended using high strength proppants or proppant concentrations of at least 1 lbm/ft^2 to overcome any uncertainties associated with these formation properties. To simulate field conditions Marpaung et al. (2008) developed a dynamic fracture conductivity apparatus. They concluded that both increasing gas flux and using a breaker enhance cleanup efficiency. On the other hand, high polymer concentration in the fracturing fluid has a deteriorating effect on cleanup.

Awoleke et al. (2012) conducted experimental investigations on tight gas reservoir using factorial design. The factors considered for the experiments were arranged in order of decreasing influence: closure stress, temperature, nitrogen rate, polymer loading, proppant concentration, and the presence of a breaker. The work is also an application of advanced statistical analysis in the field of petroleum engineering. Rivers et al. (2012) studied coated and uncoated proppant pack conductivity using consolidated sandstone cores. Resin from coated proppants reduces the pore throat diameter as a result of proppant pack consolidation, so uncoated proppants demonstrated higher conductivity performance than coated proppant. To consider geochemical reactions of formation and proppant diagenesis, a new dynamic system was set up by Aven et al. (2013). The dynamic flow setup was run for six months to simulate real field diagenesis. The results emphasized the importance of considering proppant compatibility with geochemical reactions before designing fracturing treatments.

Barree et al. (2016) developed a generic conductivity model for white sand data to Stim-Lab in 2012. Later the extension was proposed to Stim-Lab in 2013 including resin-coated and ceramic proppants. The model is effectively applicable for almost all types of proppants including sand, resin-coated, and ceramic proppants. The main objective of this work was to compare the performance of different proppants, and also predict the performance of unknown or non-standard proppant materials. The factors considered for the model are closure stress, proppant concentration, proppant properties, and rock properties. It is also possible to determine the impact

of particle size and size distribution on proppant conductivity with appropriate application of the models. Although the models provide generic baseline conductivity for the known samples but the uncertainty associated with the models are not quantified. Therefore, the accepted ranges of prediction for a standard material is yet to be defined. Thus, the accuracy of the characterization of a non-standard material also becomes questionable.

Very few statistical approaches to model fracture conductivity are found in the literature so far. However, Awoleke et al. (2016) developed one empirical and one semi-empirical model of fracture conductivity using statistical tools. They used fractional factorial design to define the number of experiments and later employed regression analysis to model fracture conductivity. For the semi-empirical model, they performed dimensional analysis to derive two dimensionless groups. They developed the relationship between the two dimensionless groups by using Buckingham Pi Theorem (1914). Later non-linear regression analysis was employed to model fracture conductivity. They considered flow back rate, reservoir temperature, polymer loading, the presence of a breaker, closure stress, and proppant concentration as the affecting factors on fracture conductivity. Kainer et al. (2017) collected four shale samples from Barnett, Fayetteville, Marcellus, and Eagle Ford to analyze their fracture conductivity behaviors. Studied parameters for the different shale formation are: closure stress, Young's modulus, Poisson's ratio, brinnel hardness number, mineralogy, proppant size, proppant concentration, and fracture surface. They employed multiple linear regression analysis to develop models of fracture conductivity for each test sample. Influential parameters for each sample were ranked by their corresponding p-values which provide an order of significance among them. A more general conductivity model was proposed which considers all four types of shale samples. They concluded that closure stress and proppant concentration hold greater significance on fracture conductivity than rock mechanical properties and fracture surface characteristics. Young's Modulus is also found to be one of the most important parameters that has a positive correlation to fracture conductivity at each stress level.

2.4 Source of the Experimental Data of This Study

This study includes around 2700 experimental data from fifteen research papers and one thesis dissertation on hydraulic fracturing. The list and the main objectives of these studies are given below:

- Cooke (1975) – To evaluate the effect of fracturing fluid on fracture conductivity.
- Vlis et al. (1975) – To determine criteria for proppant admittance and placement and to develop empirical correlations to estimate fracture conductivity.
- Cutler et al. (1985) – To study the performance of different proppant materials at intermediate and high closure stress.
- McDaniel (1986) – To investigate the effects of time, closure stress and temperature on fracture conductivity for various proppants.
- McDaniel (1987) – To evaluate realistic fracture conductivity of different proppants by varying temperature.
- Much and Penny (1987) – To study long term fracture conductivity using realistic test conditions.
- Parker and McDaniel (1987) – To study and minimize the effects of gel filter cakes on fracture conductivity by achieving high proppant concentration.
- Penny (1987) – To investigate the effects of environment and fracturing fluids on long term fracture conductivity.
- Fredd et al. (2001) – To study the effects of different fracture properties on fracture conductivity for water fracturing and conventional fracturing treatments.
- Marpaung et al. (2008) – To investigate dynamic fracture conductivity and measure cleanup efficiency.
- Awoleke et al. (2012) –To investigate the effects of different factors affecting propped fracture conductivity using factorial design.
- Rivers et al. (2012) – To investigate fracture width and fracture conductivity with high proppant loading at high pressure and temperature.
- Aven et al. (2013) – To evaluate proppant damage with long term dynamic flow testing apparatus.

- Zhang (2014) – To determine unpropped and propped fracture conductivity, fracture conductivity impairment by water.
- Li et al. (2015) – To develop mathematical models to calculate proppant embedment, proppant deformation, the change in fracture aperture, and fracture conductivity.
- McGinley et al. (2015) – To evaluate fracture conductivity with samples fractured in both horizontal and vertical orientation.

The main goal of all the research mentioned above was to investigate fracture conductivity and the factors that affect it. This research also aims to collate all the experimental data from these different sources and develop a mother dataset. The objective is to use the mother dataset as the basis for the development of empirical models of hydraulic fracture conductivity. This study determines the significant affecting factors, provides assessment of uncertainty in predictions and a methodology to minimize the effect of the fact that these datasets are from disparate sources. It also represents a methodology to deal with missing data in the field of petroleum engineering.

Chapter 3 METHODOLOGY

3.1 Introduction

This chapter includes the data processing and methodology of this research. More precisely, it will highlight the key statistical tools used to develop fracture conductivity models. It will also represent how the data is pre-processed before employing the statistical tools. **Figure 3-1** shows the workflow of the methodology.

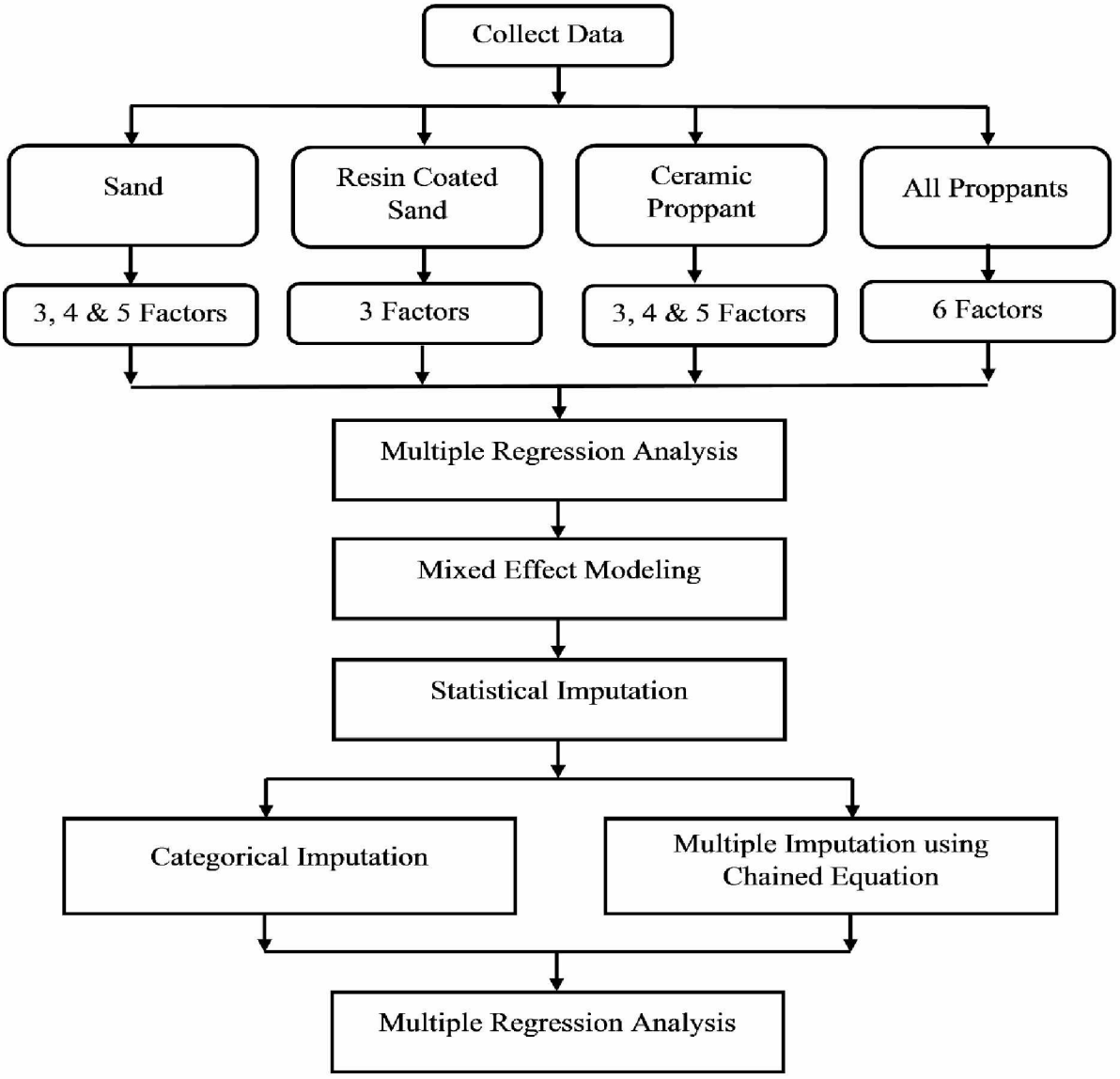

Figure 3-1: Workflow of methodology

3.2 Research Overview

To determine the behavior of fracture conductivity for different proppant types, the mother dataset initially is divided into daughter subsets according to the proppant types. Different proppant types are designated by following:

- Proppant Type-1 which is sand;
- Proppant Type-2 which is resin-coated sand;
- Proppant Type-3 which is ceramic proppants;
- All Proppant Type which is a hybrid mixture of Proppant Type-1 and Proppant Type-3.

The first step of the methodology is to employ regression analysis. To employ regression analysis, complete information of independent variables is required for each subset. In the next step, all the daughter subsets of proppant types are also divided into subsets according to the factors available from the literature. This study is limited by the availability of data collated from literature, so all the proppant types will not have an even number of fracture conductivity models. Therefore, for Proppant Type-1 and Proppant Type-3, three types of model are developed using regression analysis that are three, four and five factor models. On other hand, for Proppant Type-2, only three factor model is developed. **Table 3-1** shows the models developed by employing regression analysis along with the independent variables used to develop the models.

Table 3-1: Regression models developed in this study with the independent variables used in each model

Type of Proppant	Developed Models in this Study	Independent Variables
Proppant Type-1	Three Factor Model	Mesh Size (Proppant Particle Size), Closure Stress, and Temperature
	Four Factor Model	Mesh Size (Proppant Particle Size), Closure Stress, Proppant Concentration, and Temperature
		Closure Stress, Proppant Concentration, Temperature and Formation Hardness (BHN)
	Five Factor Model	Mesh Size (Proppant Particle Size), Closure Stress, Proppant Concentration, Temperature, and Polymer Loading
Proppant Type-2	Three Factor Model	Mesh Size (Proppant Particle Size), Closure Stress, Temperature
Proppant Type-3	Three Factor Model	Mesh Size (Proppant Particle Size), Closure Stress, Temperature
	Four Factor Model	Mesh Size (Proppant Particle Size), Closure Stress, Proppant Concentration, and Temperature
	Five Factor Model	Mesh Size (Proppant Particle Size), Closure Stress, Proppant Concentration, Temperature, and Polymer Loading
All Proppant Type	Six Factor Model	Mesh Size (Proppant Particle Size), Closure Stress, Proppant Concentration, Temperature, Polymer Loading, and Type of Proppant

In order to standardize coefficients and improve each estimate's numerical stability, a Z-score transformation is applied to all the independent variables which have quantitative magnitudes. This is also called the standard score. It indicates the number of standard deviations from the mean a data point is. Z-score transformations are applied by using the following equation:

$$Z = \frac{(X-\mu)}{\sigma} \tag{3.1}$$

In **Equation 3.1**, X, μ, and σ are the sample score, sample mean and sample standard deviation respectively. Fracture conductivity is transformed using a natural log. Qualitative independent variables are treated as categorical variables. Categorical variables take on values which are usually names, labels, groups, or types. However, this is only applicable for the regression model of All Proppant Type because only the type of proppant falls into this category.

In this study, fracture conductivity, mesh size, closure stress, proppant concentration, temperature, polymer concentration, types of proppant, and formation hardness are denoted by Fc, P_{ms}, σ_c, C_p, T, C_{pl}, t_P, and BHN respectively.

Before employing regression analysis each dataset is randomly divided into training and validation datasets by seventy percent and thirty percent respectively. Training datasets are used to develop models and validation datasets are used to validate the models. Data points for training and validation datasets are chosen randomly so that the partition is made without any bias. Therefore, all the subsets will represent randomly chosen data points having complete information in relation to the dependent and independent variables. To compare the predictive performance of the regression models of each proppant type, it is important to validate them using a consistent dataset. Therefore, for Proppant Type-1 and Proppant Type-3, validation datasets for the five factor model are used to validate all other models of that corresponding proppant type.

Based on the results of the fixed effects models **(Table 3-1)**, more complicated models are developed using mixed effects modeling and statistical imputation. The mixed effects paradigm is used to correct for the random errors associated with the disparate sources of our data. Mixed effect modeling is carried out for Proppant Type-1 and Proppant Type-3. It is less useful to perform

further analysis on Proppant Type-2 because of an inadequate number of independent variables. The best model is chosen from three, four, and five factor models of each proppant type (which is the case for Proppant Type-1 and Proppant Type-3 in this study) to employ mixed effect modeling. This is done based on the model predictive performance of each model for a particular proppant type. For Proppant Type-1, four factor model where formation hardness is present as an independent variable is not carried out for further analysis because of an inadequate number of data points. However, this comparison is not necessary for All Proppant Type as only one model is available for further analysis of this proppant type. After the comparison based on the model predictive performance, the five factor model becomes the optimal model for each proppant type, and this will be further discussed later in section 4.1.6.

Statistical imputation is used to account for missing values in some of the datasets. The primary aim of employing statistical imputation is to predict missing data and develop an expanded dataset for modeling fracture conductivity. Therefore, statistical imputation is employed for Proppant Type-1 and Proppant Type-3. Two types of statistical imputation tools are used to impute the missing values: categorical imputation and multiple imputation using chained equations.

Both mixed effect modeling and statistical imputation are performed using coefficient estimates from the best fixed effects model for each proppant type. **Table 3-2** shows the models developed from mixed effect modeling and statistical imputation.

Table 3-2: Models developed in this study using mixed effect modeling and statistical imputation

Method	Types of Proppant	Developed Models in this study
Mixed Effect Modeling	Proppant Type-1	Five Factor Model
	Proppant Type-3	Five Factor Model
Categorical Imputation	Proppant Type-1	Five Factor Model
	Proppant Type-3	Five Factor Model
Multiple Imputation using Chained Equations	Proppant Type-1	Five Factor Model
	Proppant Type-3	Five Factor Model

3.2 Regression Analysis

Regression analysis is a statistical approach to determine the relationship between a response variable and predictors. It shows how a response variable changes in association to the predictors. It also provides future predictions of the response variable on predictors and measures the importance of the predictors on the response variable. Multiple regression analysis is an extension of simple regression analysis. Simple regression analysis finds the relationship between one response variable and one predictor. On the other hand, multiple regression analysis estimates the relationship when there are two or more predictors involved. The response variable is also known as the dependent variable and predictors are known as independent or regressor variables. In this study, natural log of fracture conductivity is the response variable or dependent variable and other factors affecting conductivity are the predictors or independent variables. In this study, multiple regression analysis is performed using the lm() function in R (R Core Team, 2017). Linear regression will be described in general followed by a special case of linear regression called polynomial regression.

3.2.1 Linear Regression

Linear regression is the approach to developing the relationship between the dependent variables and the independent variables by fitting a linear equation. Linear regression can be used to estimate the following two types of effects:

- Main Effects
- Interaction Effects

3.2.1.1 Main Effects

A main effect is the effect of one independent variable on the dependent variable if all other independent variables in the model are held at a fixed value. The number of main effects will be the same as the number of independent variables. The basic equation of a linear regression using main effects is given below:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ \ldots\ldots\ldots\ldots + \beta_p X_p + \in \tag{3.2}$$

In **Equation 3.2,** Y is the dependent variable, $X_1, X_2, \ldots\ldots, X_p$ are the independent variables, and $\beta_0, \beta_1, \beta_2, \ldots\ldots, \beta_p$ are the regression coefficients where p is the number of independent variables. Here, $\in$ represents the random error term.

3.2.1.2 Interaction Effects

Interaction effects are used when the main effect of one independent variable on the dependent variable changes depending on the level of another independent variable. It is very important to consider the interaction effects between the independent variables in engineering. For example, in hydraulic fracturing treatment, the effect of mesh size on fracture conductivity can also depend on the level of proppant concentration. It also makes engineering sense if closure stress and mesh size have interactions between them because the earth transmits stress on the proppant during fracture closure which can result in crushing proppants as well as a reduction in mesh size. Awoleke et al. (2012) found that at a low level of closure stress, the effect of polymer loading on fracture conductivity is larger than at the high level of closure stress. For this research, all the main and all possible two-way interaction effects are considered while developing the linear regression

model. Similar to **Equation 3.2**, the basic equation representing a regression model considering both main and two-way interaction effects is provided in **Equation 3.3**.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{1.2} X_1 X_2 + \ \ldots\ldots\ldots\ldots + \beta_{p-1.p} X_{p-1} X_p + \in \tag{3.3}$$

3.2.2 Polynomial Regression

When the relationship between the dependent variable and one of the independent variables is not likely to be linear, terms containing an integer power(s) of the predictor can be added to the base model in Equation 3.2. This is called polynomial regression. The relationship is developed as an r^{th} order polynomial where r is a natural number representing the degree of a polynomial. In some cases, polynomial regression improves predictions over linear regression. Typically, linear regression is first employed to develop a model. If the estimated effects of the predictors do not agree with prior knowledge, then polynomial regression may provide a better model. For this particular study, r=2 is used to develop polynomial regression. Polynomial regression is modeled using all the main effects and their polynomial terms. The basic equation for a polynomial regression is given below in the **Equation 3.4**. All the parameters in **Equation 3.3** and **Equation 3.4** are the same as denoted in **Equation 3.2.**

$$Y = \beta_0 + \beta_1 X_1 + \beta_{1.1} {X_1}^2 + \ \ldots\ldots\ldots . + \beta_p X_p + \beta_{p.p} {X_p}^2 + \in \tag{3.4}$$

Regression analysis is an advanced tool of statistics used to develop empirical models. **Figure 3-2** shows the major workflow of this research to perform a regression analysis. The workflow is discussed in the next few sections.

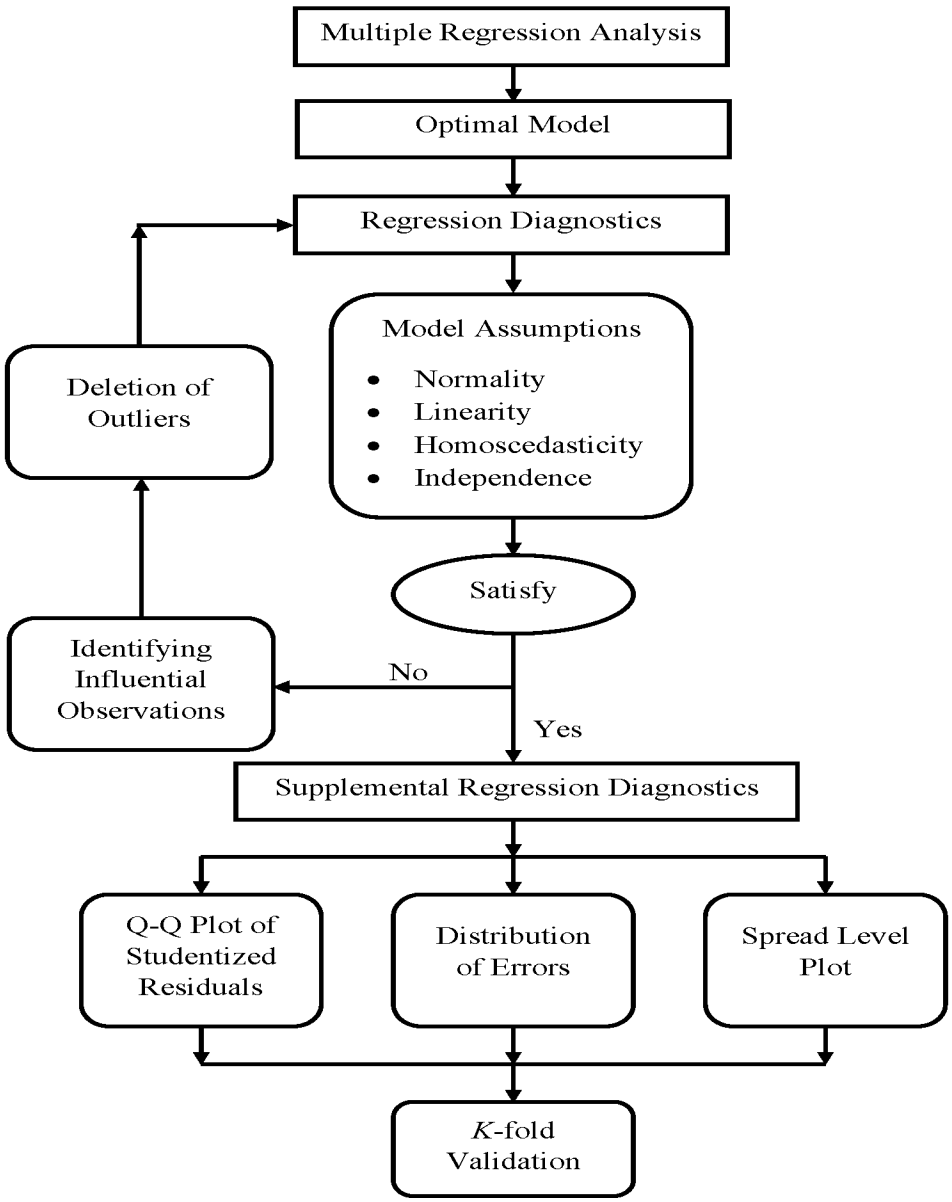

Figure 3-2: Workflow of regression analysis

3.2.3 Optimal Model

Developing a regression model with multiple independent variables, their interactions and polynomial terms can result in a potentially large number of predictors. It is normally not necessary to use all these predictors in a model. It is reasonable to choose those predictors that contribute significantly to improve the model's predictive performance. The idea is to select the best model from all possible models. To choose the best model from all possible regression models, all possible models (consisting of the models formed from every subset of the set of all possible predictors) are fit in this study. This method is called all subsets regression. It is done with the regsubsets() function of the package leaps in R (Lumley, 2017). It compares all possible models using different subsets of predictors. The analysis shows q number of best fitting models according to the criterion such as R-square, adjusted R-square, and the Mallows Cp criterion. For this study, q is 2 and either of these two models is chosen for further analysis. The criterion for selecting the optimal model is either the R-square or the adjusted R-square values. At first the criterion used is the R-square value, but if the optimal model using the R-square criterion does not include all the main effects, then adjusted R-square is used instead. If an interaction effect term or a polynomial term becomes a significant predictor after performing all subsets regression, then both the main effects of that particular term are forced into the best model regardless of their significance as main effects.

3.2.4 Regression Diagnostic

Regression diagnostics are checks performed to validate the assumptions made in the developed models. After using the lm() function in R to fit a regression model, R provides a statistical summary containing at a minimum the regression coefficients, R-square, adjusted R-square, and residual standard error. But the fitted models might not be useful for future prediction because characteristics of the data might not be consistent with assumptions made by the model. Therefore, results might be misleading, resulting in incorrect inferences about the relationship between the dependent variable and the independent variables. So it is very important to perform regression diagnostics before drawing conclusions from regression models. Regression diagnostics are designed based on the statistical assumptions in the underlying model. They can be performed with both graphical representations and statistical tests. To validate a model as an appropriate one, the data must satisfy the following major statistical assumptions:

- Linearity between the predictors and the response
- Normality in the model errors
- Homoscedasticity in the model errors
- Independence of the model errors

3.2.4.1 Linearity

This assumption states that the dependent variable has a linear relationship with the independent variables. If the dependent variable is linearly related to the independent variables then there should not be any evident relationship between the residuals and the fitted values. This assumption is validated from a graph of residuals vs. predicted natural log of fracture conductivity values. The points on the graph should be randomly and uniformly distributed around a horizontal line.

3.2.4.2 Normality

This assumption states that the dependent variable is normally distributed. If the dependent variable is normally distributed then the residuals will approximately be normally distributed with a mean of zero. This assumption is checked with a normal probability plot of standardized residuals. It is also called Normal Q-Q plot. The points on the graph should fall roughly on a straight line.

3.2.4.3 Homoscedasticity

This assumption states that the dependent variable has a constant variance among the levels of the independent variables. It can be evaluated from a graph of the square root of the absolute value of the standardized residuals vs. the predicted natural log of fracture conductivity values. This is also called scale location plot in statistics. If the constant variance assumption is met then the points on the graphs should fall randomly in a band centered on a horizontal line.

3.2.4.4 Independence

This assumption says that the dependent variable values should be independent of each other. There is no general graphical check to validate this assumption, since dependence structure depends more on the nature of data or how the data is collected. All the actual fracture conductivity values in this study have come from experiments done by varying the independent variables. The independent variables were the only changing factors controlling fracture conductivity, since all other factors were presumably controlled or randomized. Therefore, it can be assumed that fracture

conductivity values are reasonably independent of one another so that the assumption of independence is met.

3.2.5 Identifying Influential Observations

To perform regression analysis, it is frequently assumed that the response variables are normally distributed. If necessary, mathematical operations can be applied to remove non-normality like the log-transformation, Z-score transformation. In spite of such transformations, there could be some influential observations that impact the model predictive performance. Removing such observations when justified by non-statistical considerations, can result in improvement in the model performance. Cook (1977) introduced a method to find out the influential observations on data, which is known as Cook's distance. Montgomery et al. (2011) used the following equation to determine Cook's distance:

$$D_i = \frac{r_i}{p}\frac{h_{ii}}{(1-h_{ii})} \tag{3.5}$$

Where Cook's distance is denoted by D_i. Studentized residuals and leverage scores for i-th data unit are denoted by r_i and h_{ii} respectively. Studentized residuals and leverage scores are defined by **Equation 3.6** and **Equation 3.7** respectively.

$$r_i = \frac{e_i}{\sqrt{\sigma^2(1-h_{ii})}} \tag{3.6}$$

$$h_{ii} = X_i(X^TX)^{-1}{X_i}^T \tag{3.7}$$

One proposed threshold to determine an influential observation is defined by $D_i > 4/(n-k-1)$ where n is the sample size and k is the number of predictor variables (Kabacoff, 2015). It is important to mention that deleting an outlier or influential observation should necessarily improve a model's prediction. So this method is applied step by step until the R-square or adjusted R-

square shows deterioration or comparatively less improvement than the previous step. But it is not justified to throw out data points just because doing so results in a better model fit. Accordingly, the deletion of outliers in this work can be justified because of the following reasons:

- The data points are the results of experiments that have been conducted since 1975. A lot of advancement has been made so far to measure fracture conductivity as accurately as possible by simulating actual reservoir conditions. So the accuracy of the fracture conductivity measurements in the earlier research is questionable. For an example, Penny (1987) introduced the use of sandstone instead of steel platens to simulate the scenario of proppant embedment. So there is a good chance that the fracture conductivity was overestimated in the experiments done before 1987.
- The primary researchers of the data sources are also not available to verify whether they had encountered any issues during any particular measurement or not.
- The data points are manually collected from the graphs and the tables of existing published literature. So there could be uncertainty in accuracy of the collected data points.

Therefore, it seems reasonable to assume that the outliers or influential observations are results of inaccurate values of the variables.

3.2.6 *K*-Fold Validation

K-fold validation is employed for model evaluation. It provides an indication of how well the model will predict given new dataset(s). To employ *K* -fold validation the entire sample of a data set is divided into *K* number of folds or sub-samples. (*K* -1) sub-samples are treated as training data and the rest as validation data. The model is fit on the training data and later the prediction is validated using the validation data. Each of the *K* subsamples serves as a validation or test data and the combined observations from remaining *K*-1 subsamples serve as the training data. Kabacoff (2015) presented a code using the bootstrap package in R. Kabacoff's function is used to employ *K*-fold validation in this study. *K* is set to be 10. The choice of *K* is arbitrary. The function provides an average summary of original R-square and 10-fold cross validated R-square. The original R-square is based on the sample and the cross validated R-square represents how well the model will fit with new data. The change in R-square represents the quality of a model.

3.2.7 Relative Weights of the Predictors

Johnson (2000) proposed a method to estimate the relative weight of the predictors of a regression model. This method takes into account the correlation of the variables in data. This method is distinct from testing which predictors are significant and it conditions on which predictors are included in the model. Hence it is an assistance to provide a ranking of the significant variables that are determined from the variable selection methods to belong in the optimal model. This method is employed on the best regression models of Proppant Type-1 and Proppant Type-3 to develop a ranking of the predictors. The results allow us to decide relative importance of the predictors. The R code used for this method is presented by Kabacoff (2015), which is a modification from Johnson (2000).

3.3 Mixed Effect Model

Mixed effect model is a statistical tool for modeling the relationship between the dependent variable and independent variables when both fixed and random effects are important. These two types of effect will be discussed briefly later in this section. When there are repetitive measurements/observations of the same units, or disparate sources of data, or clustered data, an unwanted correlation in the response might arise because of not being independent within a data source. In mixed effect modeling, along with the fixed predictors one or more categorical variables are used representing the levels of each random effect. The basic equation for a mixed effect model is given below in **Equation 3.8.**

$$Y = X\beta + Zu + \in \tag{3.8}$$

Fixed effects use categorical predictors for which inferences are to be confined to the specific levels included in the study. In **Equation 3.8**, $X\beta$ term denotes the fixed effects. On the other hand, random effects are used to refer to the effects of different levels of a categorical predictor including levels not included in the study. The random effects assumption states that the data collected for analysis are from a finite number of different populations. The term Zu in **Equation 3.8** is called the random effect of a mixed model.

This research is based on multiple sources of hydraulic fracture conductivity data. The data are obtained from repetitive measurements of factors affecting fracture conductivity. Therefore, mixed effect modeling is employed as an extension of the previous work. Disparate sources are labeled as different levels of a categorical variable. The most significant predictors included in the optimal models from regression analysis are used in this method for each proppant type. Mixed models are fit using the linear mixed effect modeling function lmer() from the package lme4 in R (Bates et al., 2015).

3.4 Statistical Imputation

Statistical imputation refers to the methods of imputing the missing values of a sample. Instead of ignoring the values that cannot be acquired or recovered, imputation predicts those values. The imputed values contribute to a larger set of data to develop the relationship between the dependent variable and independent variables. There are several methods to impute missing values. But before choosing a method, it is important to understand the nature of the missing data. Little and Rubin (2002) proposed the classification missing data with the following:

- Missing Completely at Random **(MCAR):** Probability of a value being missing is not related to any other variables of an observation or the value itself.
- Missing at Random **(MAR):** Probability of a value being missing depends on the other observed variables but not on the value itself.
- Missing Not at Random **(MNAR):** Probability of a value being missing depends on the variables which are missing.

All the variables in this study were observed by the researchers and then reported in the literature. Since the reporting of a variable probably has very little to do with that or values of other variables, so MCAR or MAR can be assumed. For this study only the missing values of polymer loading are predicted using statistical imputation. There are two major reasons for choosing polymer loading to be imputed:

- In most of the previous hydraulic fracture conductivity experiments, polymer loading was either kept at a certain level or varied in a small range, so there are fewer chances to impute the missing values inaccurately.

- Other independent variables were not missing in large amounts, so those few data points can be ignored for further analysis.

3.4.1 Categorical Imputation

If the missing mechanism is MCAR, then there is no way to impute those from the other observations. Thus, it entirely depends on the independent researchers. The incomplete mother dataset provides a reason to believe that polymer loading for each source was maintained at a certain level for the experiments. For an example, McDaniel (1987) is one of the sources for this study. He conducted more than one hundred hydraulic fracture conductivity tests with no value of polymer loading presented in that literature. But it has been a common practice from the past to maintain at least a low level of polymer loading while conducting fracture conductivity experiments. Therefore, missing values of polymer loading can be assumed to be categorical or qualitative variable where each source of data represents a constant level of polymer loading. Then regression analysis is employed with this categorical polymer loading variable. However, the interpretation of a coefficient associated with a categorical variable is different from the regular interpretation of a regression coefficient associated with a continuous variable. It depends on how the reference level is set up for the categorical variables. For simple linear regression in R, the intercept is equal to the mean response for one of the levels of categorical variables. That level is defined as the reference level. The coefficient for one of the other levels of categorical variables represents the difference in mean response of that level and the reference level.

3.4.2 Multivariate Imputation Using Chained Equations (MICE)

Multiple imputation is a statistical technique to investigate a dataset with missing values. Rubin (1987) proposed the theory of multiple imputation in the context of nonresponse in surveys. It replaces each missing values with two or more plausible values. Plausible values are a random selection from an observed distribution that are similar to their parent distribution. Multiple imputation using chained equations is an approach to impute values using the algorithm of Gibbs sampling. Gibbs sampling is an MCMC (Markov Chain Monte Carlo) technique that was proposed by Geman, S. and Geman, D. (1984). It predicts the missing values from all other available variables in the dataset by generating posterior samples of each variable or block of variables. The technique involves estimating by simulation the expectation of a statistic in a complex model and does successive random selections which form a Markov Chain (Gilks, 2005). A Markov chain

refers to a set of states in which the process starts in one of those states and moves successively from one state to another (Grinstead and Snell, 1997). The probability of the next move does not depend on the current state. The process continues until the plausible values are achieved.

In this study, multivariate imputation using chained equations is done using mice package in R (Buuren and Oudshroon, 2011). The approach includes three steps which are listed in Kabacoff (2015): imputation, analysis of results, and pooling. **Figure 3-3** represents the steps in applying multiple imputation. Missing values are imputed using the mice() function in R which returns *m* number of completed datasets. (By default, *m*=5 in R.) By default, R imputes the missing values predictive mean matching (PMM). PMM replaces the missing values that are similar to the real values (Allison, 2015). In other words, the distribution of the imputed values are similar to the distribution of the real values. The next step is to analyze the *m* completed datasets using the with() function in R. The interest of this study is to develop models using linear regression. Therefore, *m* number of linear regression models are fit for m number of imputed datasets using the with() function. Linear regression is employed using the lm() function. And finally, the results are pooled using the pool() function. By default, the pool() function provides a summary with the means of all the coefficients.

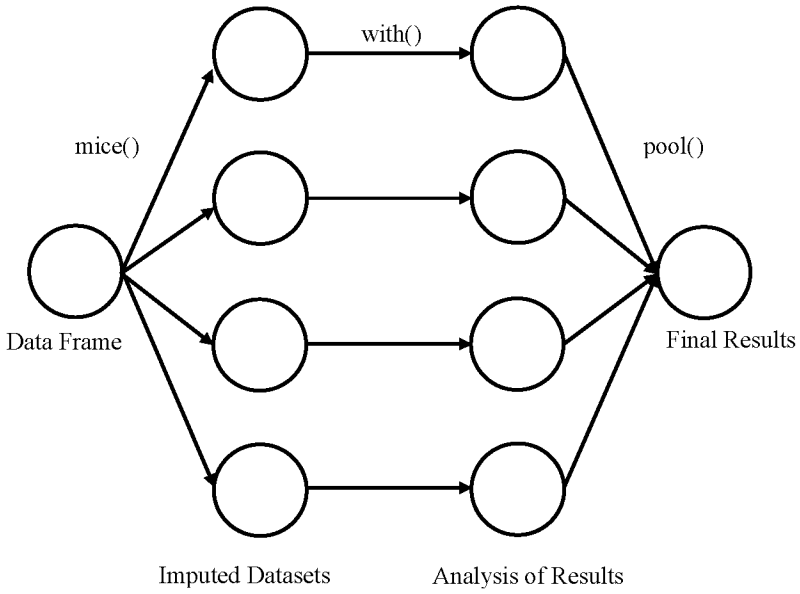

Figure 3-3: Steps to employ multiple imputation using chained equations using mice in R (Kabacoff, 2015)

This method is employed assuming the missing mechanism to be MAR. However, the fracture conductivity models are developed by altering the pooling step of mice package in R because the primary interest of this study is not to find the missing values, but deriving models of fracture conductivity. This study iterates for m =2000, so two thousand sample runs are conducted, which return two thousand imputed datasets. Each dataset provides one regression model of fracture conductivity. Instead of using all the sample runs, only the last thousand sample runs are analyzed and the median of each coefficient is reported as the final results. This is done in order to avoid any kind of convergence issue of the initial sample runs.

Finally, the predictive performances of all the models developed using different statistical approaches are compared using validation data and new data from case histories. In order to use the models developed in this research, the new dataset of independent variables must be standardized using the mean and standard deviation provided in **Appendix B**. For each model, a range is provided for each independent variable in **Appendix B**. If a new predictor data does fall outside the range, then the model's prediction is doubtful because linearity tends to break down at some point outside the range. Therefore, it is recommended to use a model if the new predictor data falls in between its corresponding range of independent variables.

Chapter 4 RESULTS AND DISCUSSIONS

4.1 Fixed Effect Model

The main purpose of this study is to develop models to predict fracture conductivity and also to report the significant predictors. Statistically significant predictors are identified by their p-values. A p-value below 0.05 indicates a predictor to be statistically significant. However, regardless of the significance, all the predictors reported from all subsets regression are chosen for further analysis.

4.1.1 Proppant Type-1

4.1.1.1 Three Factor Model

Around 750 data points are used to develop this model. The first linear regression model has an R-square of 0.4732 and an adjusted R-square of 0.4689. **Figure 4-1** represents the optimal model where all the possible main effect and interaction effect terms are in the X-axis. The colored blocks in each row represent the variables of a model at its corresponding adjusted R-square. Starting from the bottom, the first row represents a model with intercept and the interaction term of mesh size and closure stress that has a very low adjusted R-square of 0.24. In the top two rows, the black blocks represent the predictors of the best two models at highest adjusted R-square of 0.47. Starting from the top, the first row does not include the interaction of mesh size and temperature. Afterwards, in the second row, the main effect of temperature is not included in the best model as a predictor. The predictors from the first row (starting from the top) are chosen to be the optimal model or best model because this study is more interested in developing models that contain all the main effects at the minimum. Therefore, the interaction of mesh size and temperature is not included as a predictor in the optimal model.

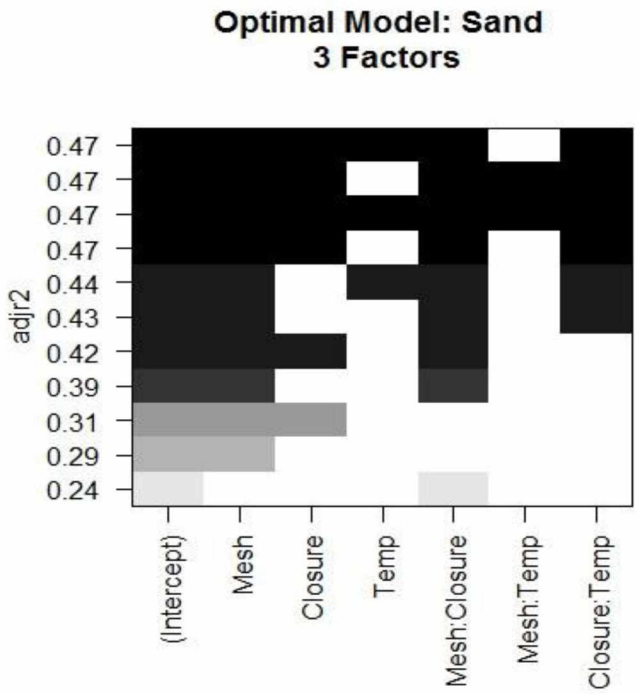

Figure 4-1: Optimal model from all subsets regression for three factor model of Proppant Type-1

To improve the model, Cook's distance is employed to determine the influential observations. **Figure 4-2** shows the Cook's distance plot before deletion (left) and after deletion (right) of these influential observations. The blue line represents the threshold discussed in section 3.2.5. Any data points above the threshold are believed to have significant influence on the regression model. The left plot of **Figure 4-2** shows that the 646th observation has a higher Cook's distance and it might have a notable impact on the intercept and slopes of the regression model. The 646th observation is deleted from the dataset and then another regression model is developed using the new dataset. The changes in R-square and adjusted R-square are observed. Both R-square and adjusted R-square show improvement after the first attempt of deletion. The process of deleting influential observations and developing a regression model on the new dataset is continued until the current regression model shows improvements from the past model. In the right plot of **Figure 4-2**, there are certain data points that are above the threshold. But deleting the 693th observation reported in the plot does not provide a significant improvement in the model. Finally, 23 influential data points are deleted. The R-square and the adjusted R-square of this final model are 0.5205 and 0.5171

respectively. After deleting these influential observations the mean squared error (MSE) is reduced from 2.28 to 1.83.

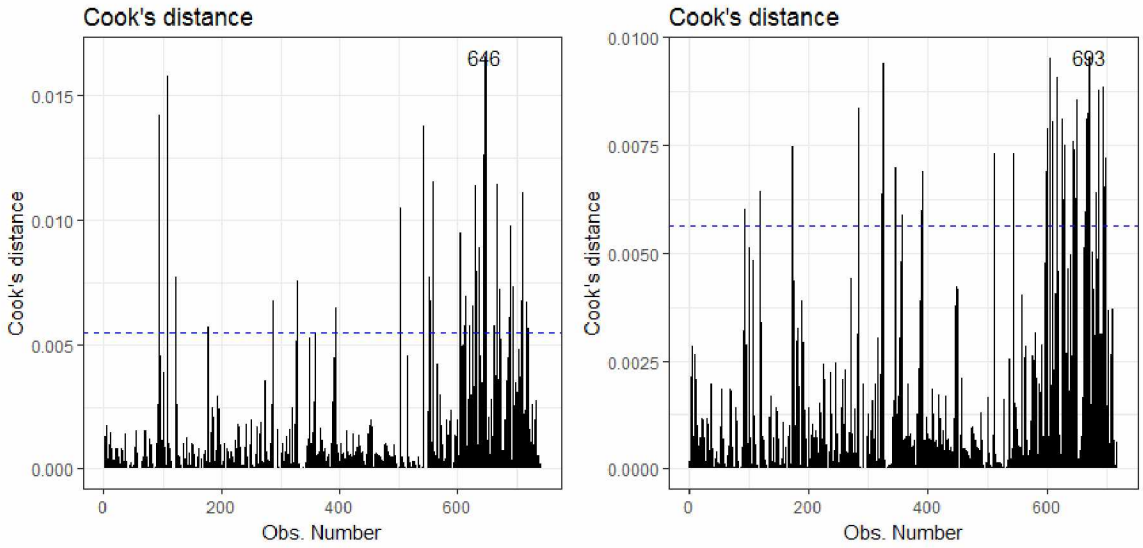

Figure 4-2: Cook's distance plot for three factor model of Proppant Type-1 (Left-before deletion, Right-after deletion)

Figure 4-3 is the regression diagnostic before deleting the influential observations and it consists of three plots for the validation of assumptions of linearity, normality and homoscedasticity. In the residual vs. fitted plot (upper left), there is some evidence of a curved shape formed by the residuals cloud around the horizontal blue line. However, there is strong evidence of straight vertical lines around the blue line which suggests that the predictors do not have many distinct levels. Standardized residuals are used to generate the normal Q-Q plot and scale-location plot. In the normal Q-Q plot (upper right), there are certain points at the top and bottom which do not fall on the line (blue line) and the distribution is left skewed. In the scale-location plot (bottom), there is a downward bend developed by the points. Therefore, the constant variance assumption is perhaps not met. Overall, regression diagnostics from **Figure 4-3** suggests violation of the model assumptions.

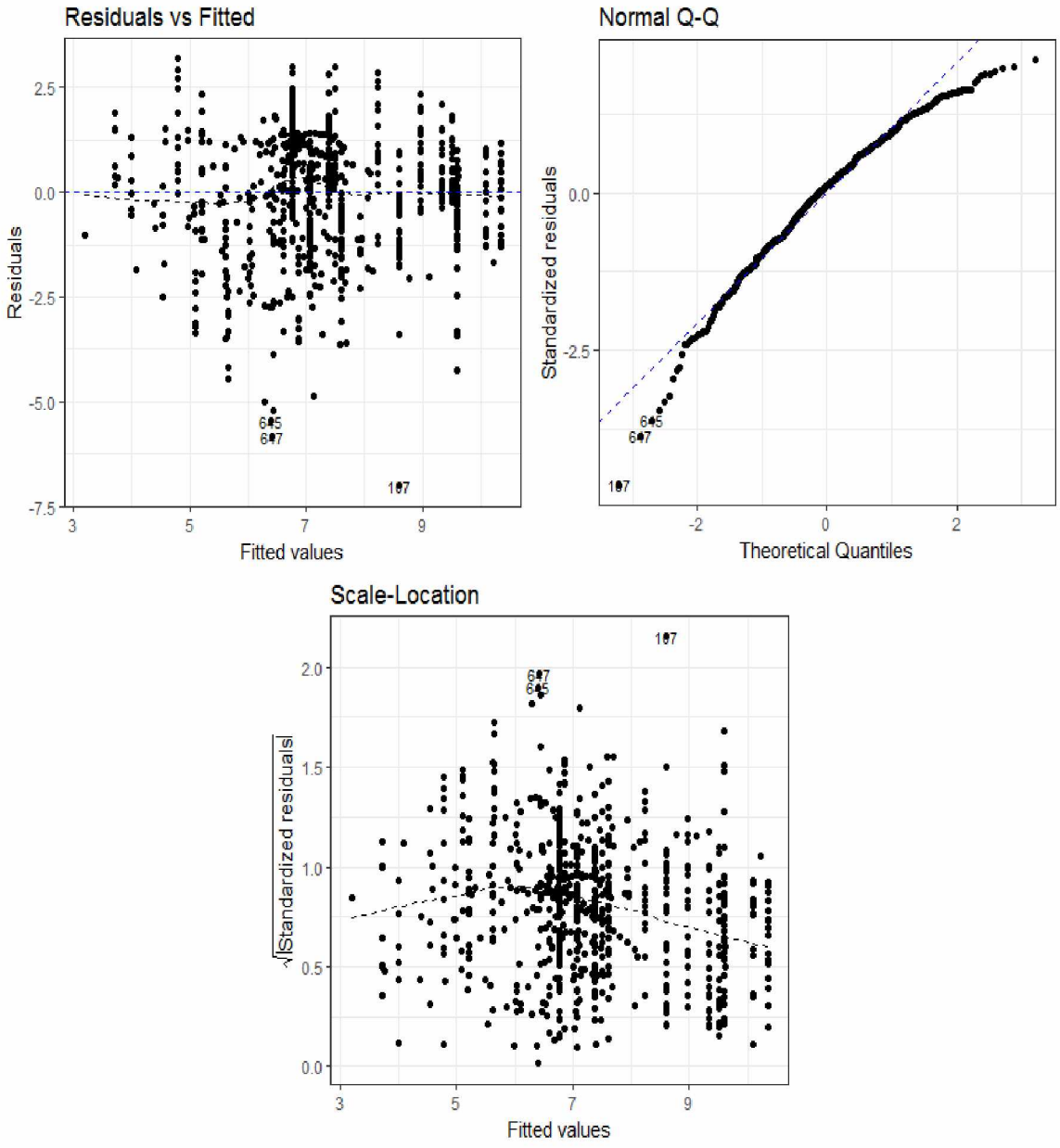

Figure 4-3: Regression diagnostic before deleting influential observations as the validation of the model assumptions for three factor model of Proppant Type-1, (Upper left-Linearity, Upper right-Normality, Bottom-Homoscedasticity)

Figure 4-4 is the regression diagnostics after deleting the influential observations. It shows little improvement from the very first model. In the residual vs. fitted plot (upper left), the residuals got more scattered around the horizontal line. In the normal Q-Q plot (upper right), only few data points at the top don't fall on the line and skewness is slightly reduced. In the scale-location plot

(bottom), the residuals are more uniform than the first model but there is still a downward bend by the data points. Overall, the model does not meet all the assumptions adequately even after deleting the influential observations.

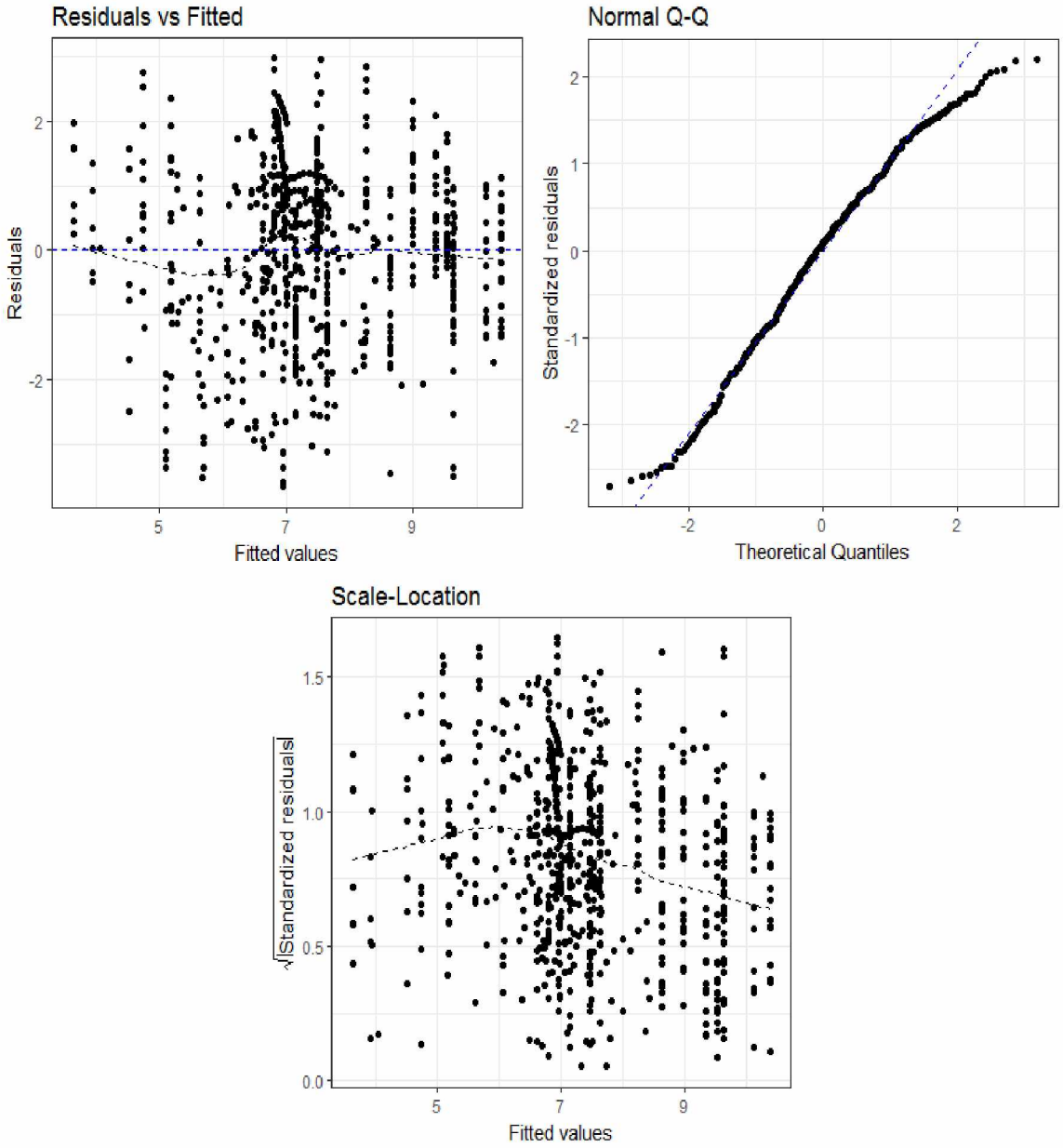

Figure 4-4: Regression diagnostic after deleting influential observations as the validation of model assumptions for three factor model of Proppant Type-1, (Upper left-Linearity, Upper right-Normality, Bottom-Homoscedasticity)

Normality and homoscedasticity assumptions are further assessed by the studentized residuals (**Equation 3.6**). Studentized residual is one scaled residual that is more effective in determining the outliers in multiple regression analysis. To assess the normality assumption more accurately, the studentized residuals are plotted against a t distribution with n-k-1 degrees of freedom where n is the sample size and k is the number of regression parameters. This plot is generated by the qqPlot() function in R. **Figure 4-5** shows the Q-Q plot for studentized residuals which is similar to the normal Q-Q plot of **Figure 4-4**. The red envelope indicates a 95% confidence interval. There are certain data points that are outside the confidence interval envelope.

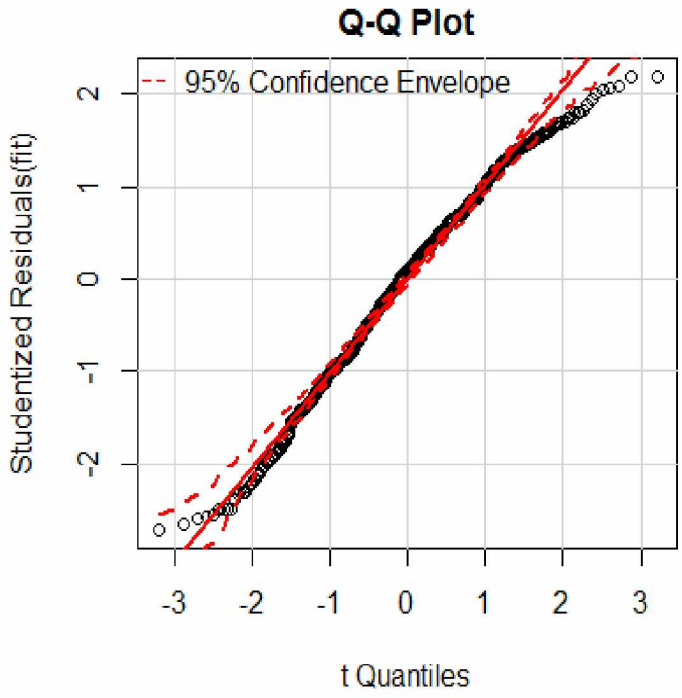

Figure 4-5: Q-Q plot of studentized residuals for three factor model of Proppant Type-1

Figure 4-6 is the histogram of studentized residuals showing the distribution of residuals, which is nearly bell shaped. It shows that the kernel density curve and normal curve are close to each other, so the normality assumption is slightly satisfactory for this model. **Figure 4-7** shows a spread level plot with the absolute studentized residuals where the red line represents the slope of the variance and the green dotted line represents the locally weighted smoothing line (LOESS

Line). The plot is generated by using the spreadLevelPlot() function in R which also provides a suggestion for the power transformation of the variables. If the value of suggested power transformation is close to 1, then it indicates that the constant variance assumption is met. In this case, the value is 1.895 and there is a non-horizontal trend indicated by the red line. Thus, it suggests a slight violation from constant variance assumption. Both the qqPlot() and the spreadLevelPlot() functions are from the package car (Fox and Weisberg, 2011).

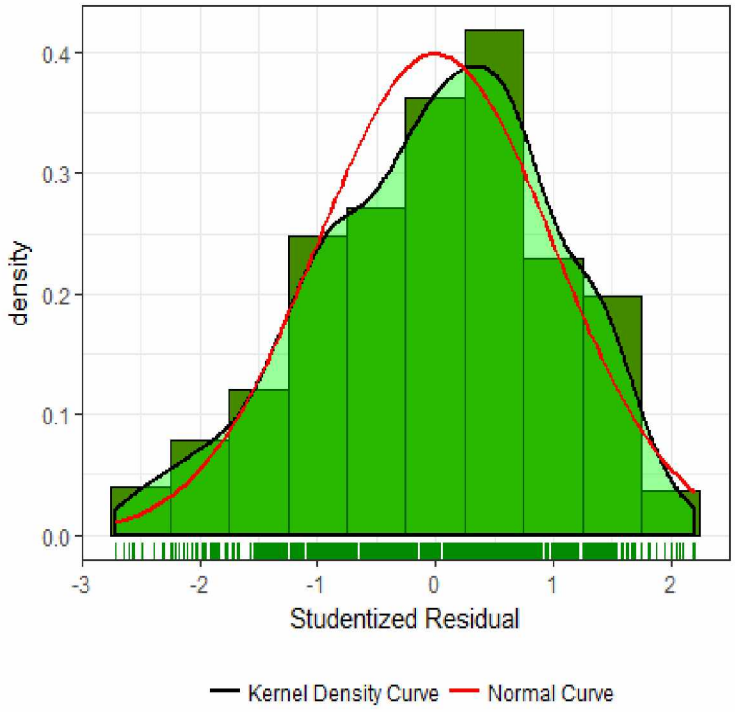

Figure 4-6: Distribution of errors for three factor of Proppant Type-1

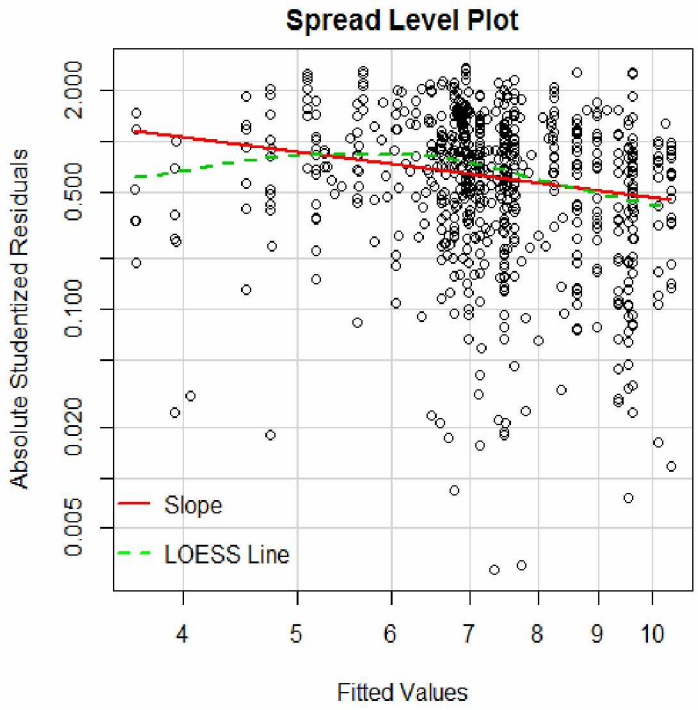

Figure 4-7: Spread level plot for three factor model of Proppant Type-1

From 10-fold cross validation, the cross validated R-square and the original R-square are reported as 0.3219 and 0.5206 respectively. Additionally, the original adjusted R-square is 0.5165 and cross validated adjusted R-square is 0.3163. The final model is proposed by **Equation 4.1**. In this equation, temperature has a p-value greater than 0.05. It should be noted that when the data was collected, all the variables (dependent and independent) were kept in consistent set of unit and later used them as input in R to develop the models. Therefore, regardless of whatever the units are for the independent variables, as long as they are consistent, their effects on fracture conductivity will be described in md-ft.

$$Fc = \exp[7.21611 + (0.75318 * P_{ms}) - (0.35250 * \sigma_c) - (0.02203 * T) - (1.04327 * P_{ms} * \sigma_c) - (0.64078 * \sigma_c * T)] \quad (4.1)$$

4.1.1.2 Four Factor Model

Two models are developed in this research using four factors for Proppant Type-1. The first four factor model includes mesh size, closure stress, proppant concentration, and temperature. Around 600 data points are used to develop this model. The first linear regression model has an R-square of 0.694 and an adjusted R-square of 0.6889. The optimal model from all subsets regression is shown in **Figure 4-8**. The optimal model does not include the interactions between mesh size and temperature or between closure stress and concentration. R-square is used as the criterion to choose the best fitting model.

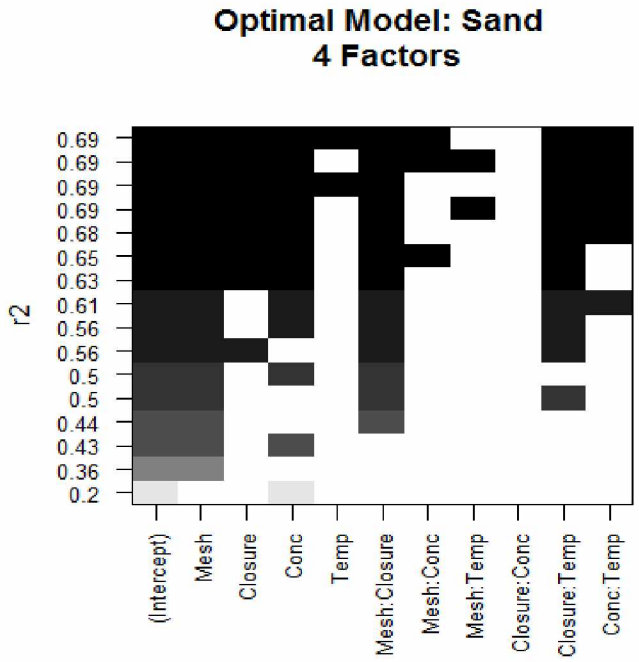

Figure 4-8: Optimal model from all subsets regression for four factor model of Proppant Type-1

Figure 4-9 shows the Cook's distance plot before deletion (left) and after deletion (right) of these influential observations. The left plot of **Figure 4-9** shows that the 24th data point is well above the blue line which indicates a clear influence of that data point on the regression model. Overall 25 influential observations are deleted to reach the final best fitting model. The right plot of **Figure**

4-9 shows that a good number of data points are above the threshold but deleting the 501th observation does not improve the regression model. The R-square and adjusted R-square of the final model are 0.7442 and 0.7407 respectively. After deleting these influential observations the mean squared error (MSE) is reduced from 1.55 to 1.18.

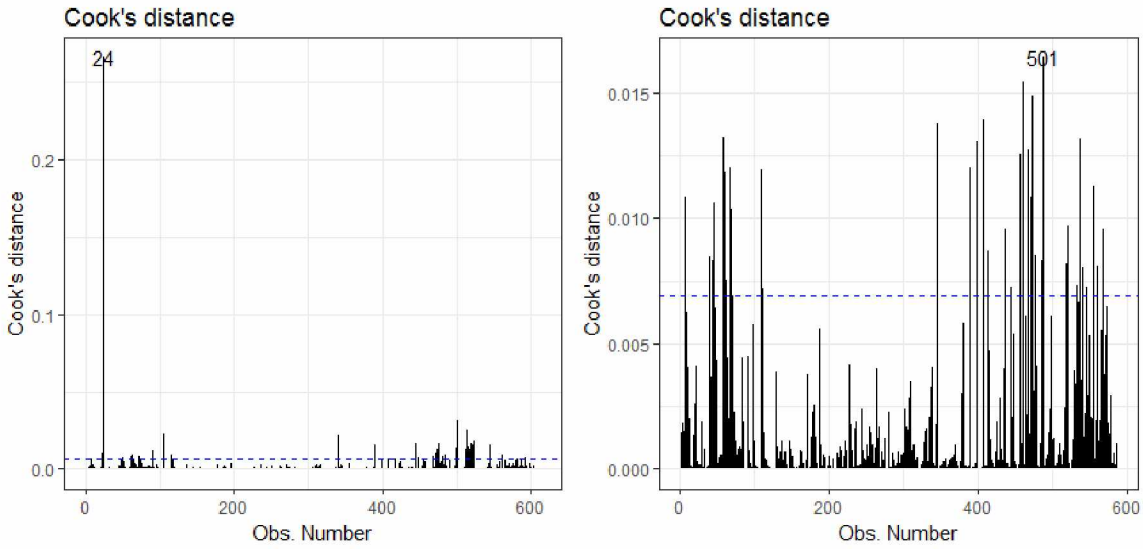

Figure 4-9: Cook's distance plot for four factor model of Proppant Type-1 (Left-before deletion, Right-after deletion)

The regression diagnostic for this model is shown in **Figure 4-10**. In the residual vs. fitted plot (upper left), there is a conical shape developed by the data points. In the normal Q-Q plot (upper right), the data points are nearly light tailed but an immense number of data points do not fall on the line. In the scale-location plot (bottom), a downward bend formed by the data points is observed. The model therefore fails to satisfy the model assumptions.

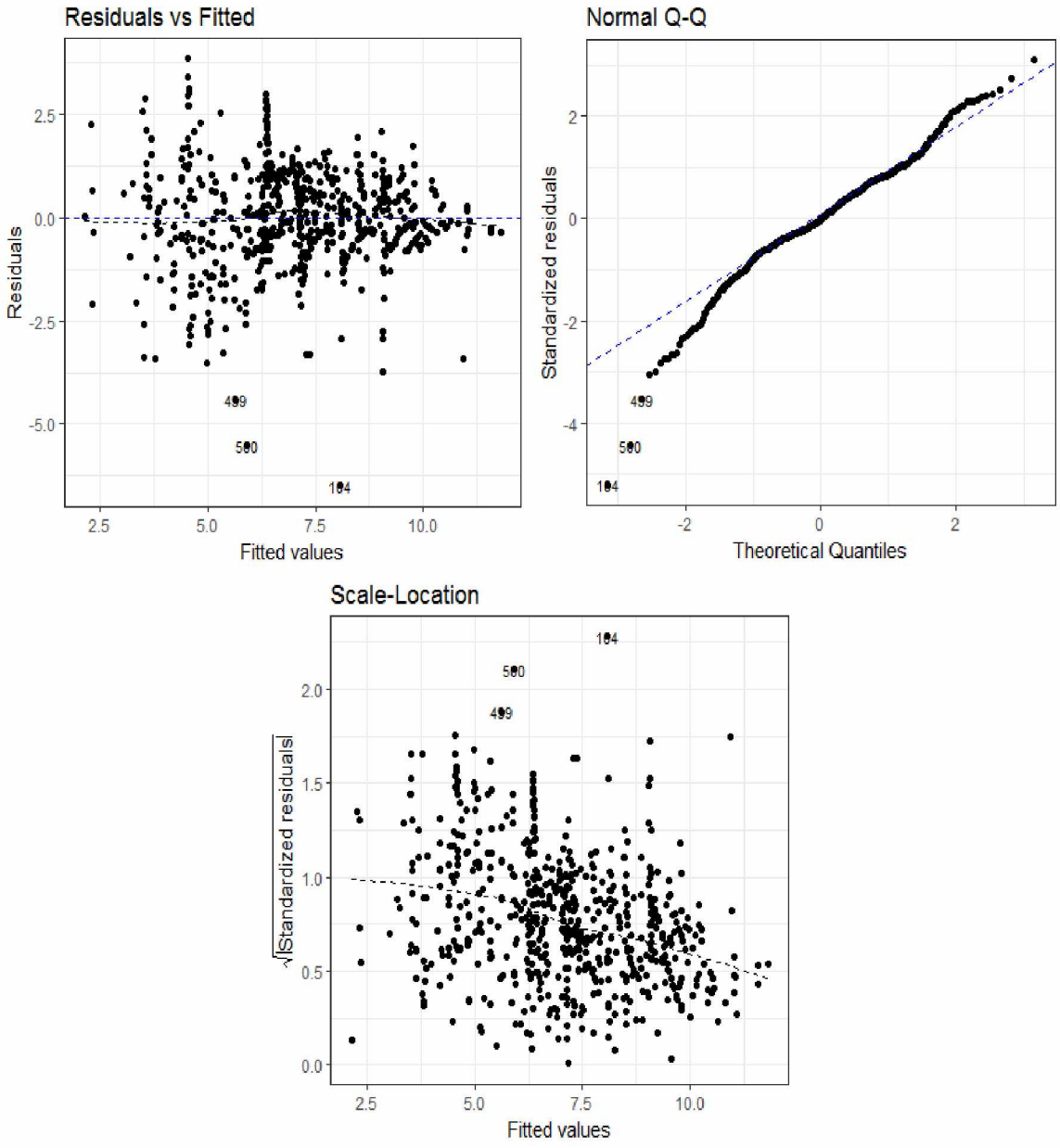

Figure 4-10: Regression diagnostic as the validation of the model assumptions for four factor model of Proppant Type-1, (Upper left-Linearity, Upper right-Normality, Bottom-Homoscedasticity)

The regression diagnostic plot after deletion in **Figure 4-11** shows that there have been no significant improvements in any of the model assumptions. Only the normal Q-Q plot (upper right) shows slight improvement from the previous normal Q-Q plot of **Figure 4-10**, but there are still a large number of data points that do not fall on the line.

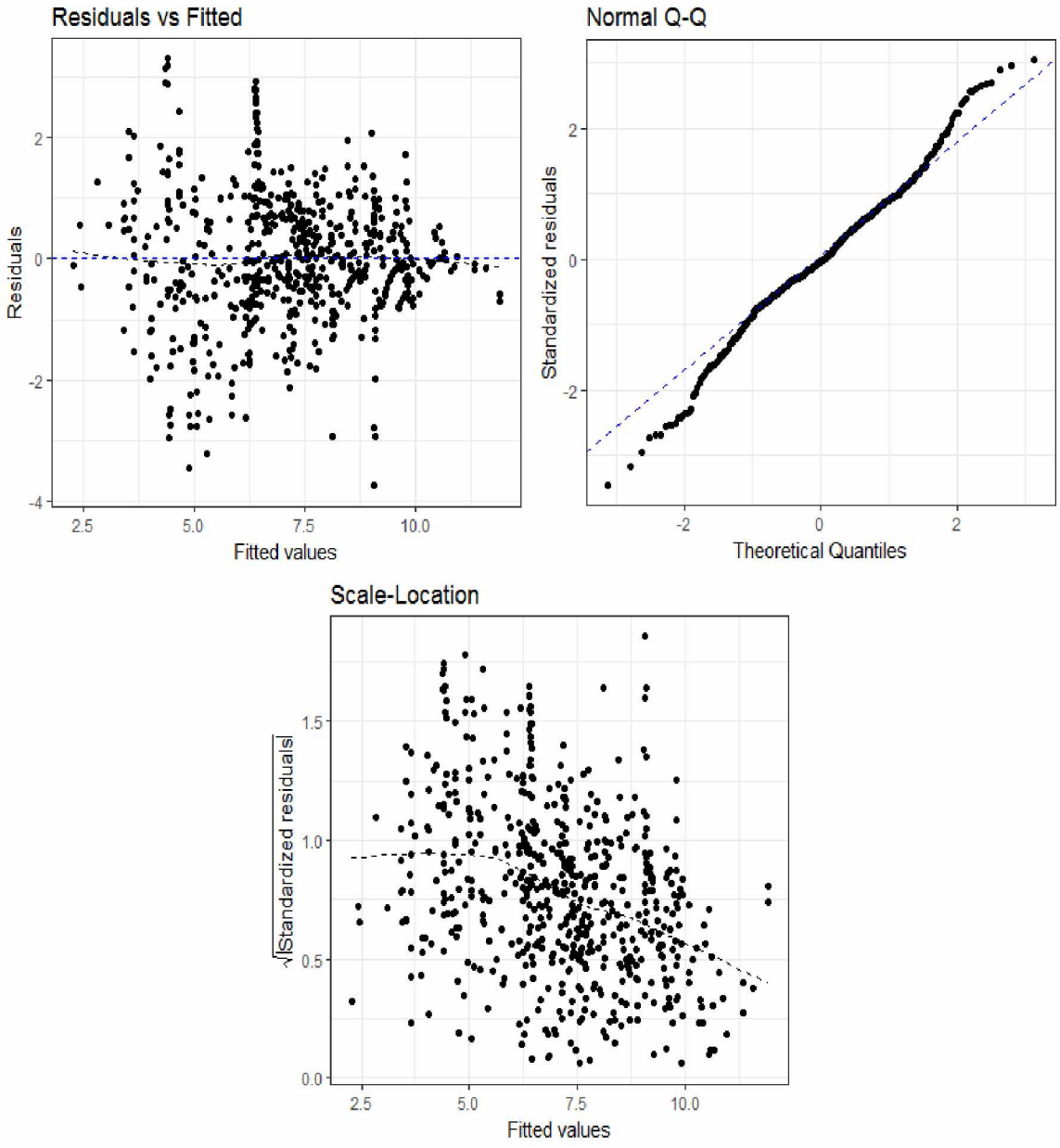

Figure 4-11: Regression diagnostic after deleting influential observations as the validation of model assumptions for four factor model of Proppant Type-1, (Upper left-Linearity, Upper right-Normality, Bottom-Homoscedasticity)

In **Figure 4-12**, the Q-Q plot of studentized residuals shows that a good number of data points are outside the 95% confidence envelope. The distribution of error plot in **Figure 4-13** shows that the normal curve and the kernel density curves are deviated from each other even though the histogram is closely bell shaped. The spread level plot in **Figure 4-14** shows a downward non-horizontal

bend indicated by the red line. The suggested power transformation for the regression model is 2.09. The regression diagnostic ascertains clear violations of the model assumptions.

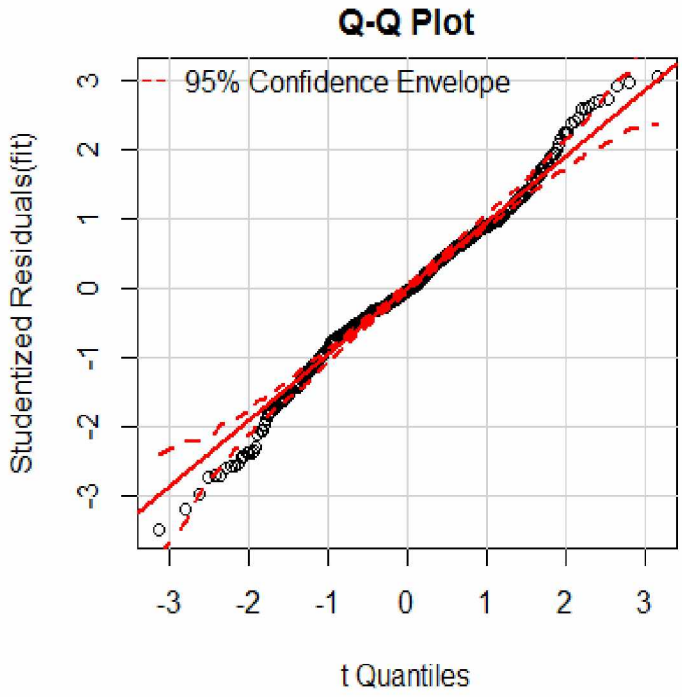

Figure 4-12: Q-Q plot of studentized residuals for four factor model of Proppant Type-1

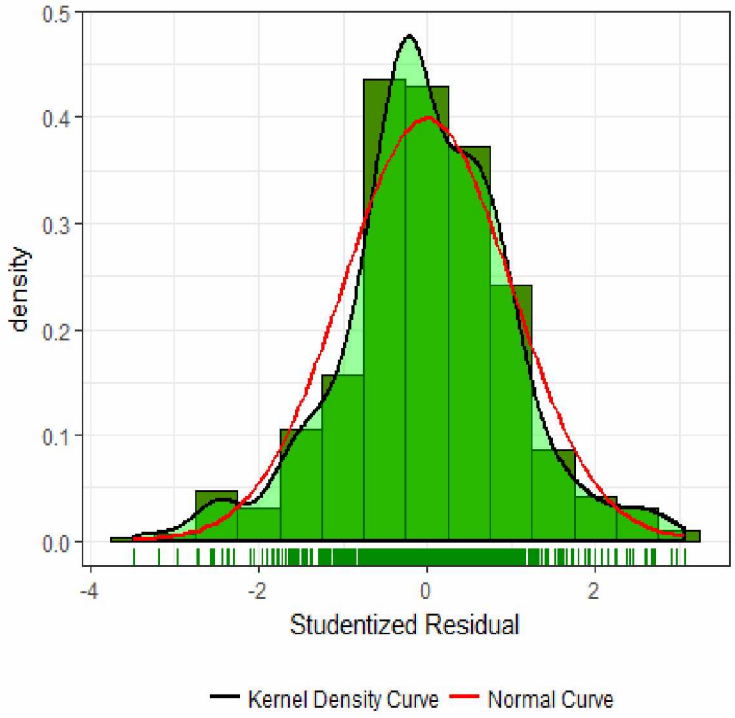

Figure 4-13: Distribution of errors for four factor model of Proppant Type-1

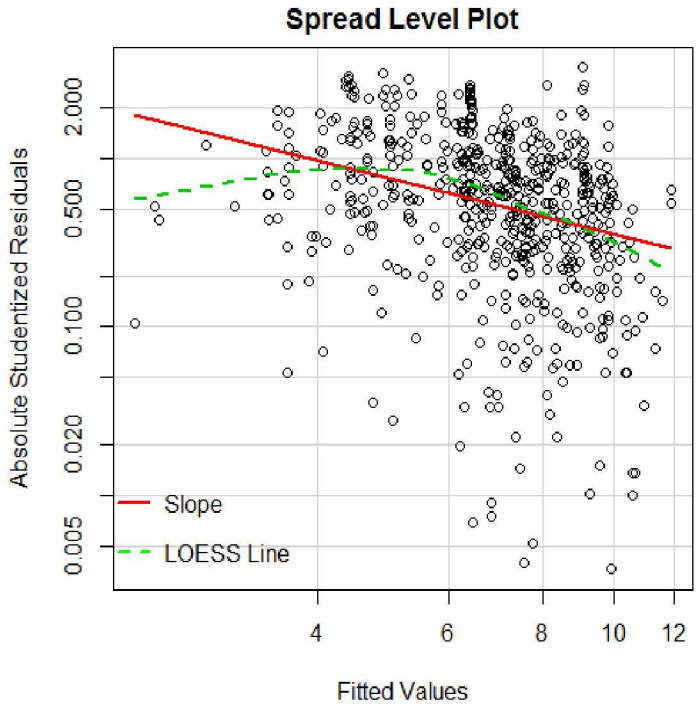

Figure 4-14: Spread level plot for four factor model of Proppant Type-1

The cross validated R-square and the original R-square are reported as 0.5106 and 0.7442 respectively. The cross validated adjusted R-square is 0.5030 and the original adjusted R-square is 0.7402. However, the model is presented in **Equation 4.2**. All the parameters have p-values lower than 0.05.

$$Fc = exp[7.20960 + (0.59221 * P_{ms}) - (0.59399 * \sigma_c) + (1.63724 * C_p) - (0.38014 * T) - (0.78710 * P_{ms} * \sigma_c) - (0.60512 * P_{ms} * C_p) - (0.69822 * \sigma_c * T) + (0.81040 * C_p * T]$$ (4.2)

The other four factor model of Proppant Type-1 includes formation hardness (BHN) as a factor instead of mesh size. It will be referred as the 2nd four factor model of Proppant Type-1 throughout this study. Around 300 data points are used to develop the model. The very first linear model developed in R provided "NA" as the regression coefficients for the interactions both between proppant concentration and temperature and between temperature and formation hardness. It necessarily suggests that the regression coefficients for these two parameters are not estimable. Collinearity might be a reason behind this, which indicates that the predictors might be linearly related. Therefore, both the predictors are excluded from the regression model. The graphical representations of the regression analysis for this model are provided in **Appendix A**. The optimal model for this model is provided in **Figure A-1**. Adjusted R-square is chosen as the criterion to select the best fitting model. The interactions between both closure stress and proppant concentration and between closure stress and temperature are left out in the optimal model. Cook's distance plots are provided in **Figure A-2**. Overall 22 data points are removed from the original dataset. After deleting these influential observations, the mean squared error (MSE) is reduced from 1.17 to 0.89. The R-square has increased from 0.5839 to 0.6433 and the adjusted R-square has increased from 0.5724 to 0.6353.

The regression diagnostic plots before and after deletion are provided in **Figure A-3** and **Figure A-4** respectively. In both regression diagnostic plots, residual vs. fitted (upper left) plot shows clear evidence of curved and funnel shape developed by the data points. The scale-location plot (bottom) has a downward bend in both of them. In the normal Q-Q plot (upper right) of **Figure A-3**, a good number of data points do fall on the line, which has even improved after deleting the

influential observations shown in **Figure A-4**. The distribution of the data points also appears to be normal in both **Figure A-3** and **Figure A-4**. In addition, both the Q-Q plot and distribution of errors plot by studentized residuals suggest to satisfy the normality assumption that are provided in **Figure A-5** and **Figure A-6** respectively. On the other hand, the spread level plot given in **Figure A-7** proves violation of the homoscedasticity assumption. The suggested power transformation is 3.86 for this case.

However, from 10-fold cross validation, the cross validated R-square and the original R-square are reported as 0.6105 and 0. 6433 respectively. The original adjusted R-square is 0.6004 and the cross validated adjusted R-square is 0.6339. Therefore, even though the regression diagnostic does not suggest the model to be adequate, it still might provide good predictions. The model is proposed by the following **Equation 4.3** where all the parameters have p-values lower than 0.05.

$$Fc = exp[8.15643 - (0.82407 * \sigma_c) + (86302 * C_p) - (0.23893 * T) - (0.20613 * BHN) + (0.24599 * \sigma_c * BHN) + (0.36996 * C_p * BHN)] \quad (4.3)$$

4.1.1.3 Five Factor Model

More than 130 data points are used to develop the model. The very first linear model developed with the dataset has an R-square of 0.9126 and an adjusted R-square of 0.903. However, the main effects of mesh size and temperature are reported as negative, which goes against the documented literature. Instead of linear regression, employing polynomial regression of 2nd order solves the problems with the flipped signs. The polynomial model has an R-square and an adjusted R-square of 0.8713 and 0.8608 respectively. **Figure 4-15** shows the optimal model where the main effect of temperature and polynomial effect of mesh size are excluded. As the polynomial effect of temperature is in the optimal model, however, so is the main effect of temperature added in the final optimal model as a predictor.

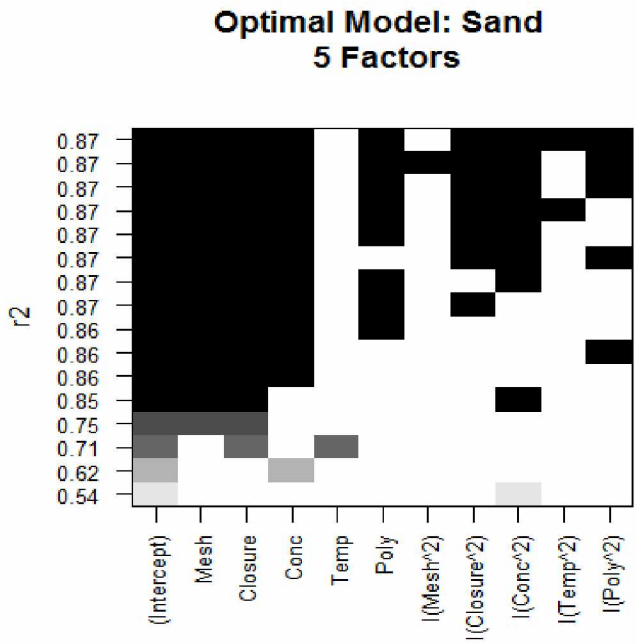

Figure 4-15: Optimal model from all subsets regression for five factor model of Proppant Type-1

Figure 4-16 shows the Cook's distance plot before deletion (left) and after deletion (right). The Cook's distance plot before deletion (left) indicates that very few data points are above the blue line which might have significant influence on the regression model. Only 3 data points have to be removed to reach the peak point where removing a data point does not influence the regression model. The final R-square and adjusted R-square for this model are 0.9056 and 0.8985 respectively. Deleting only 3 data points decreased the mean squared error (MSE) from 0.44 to 0.33.

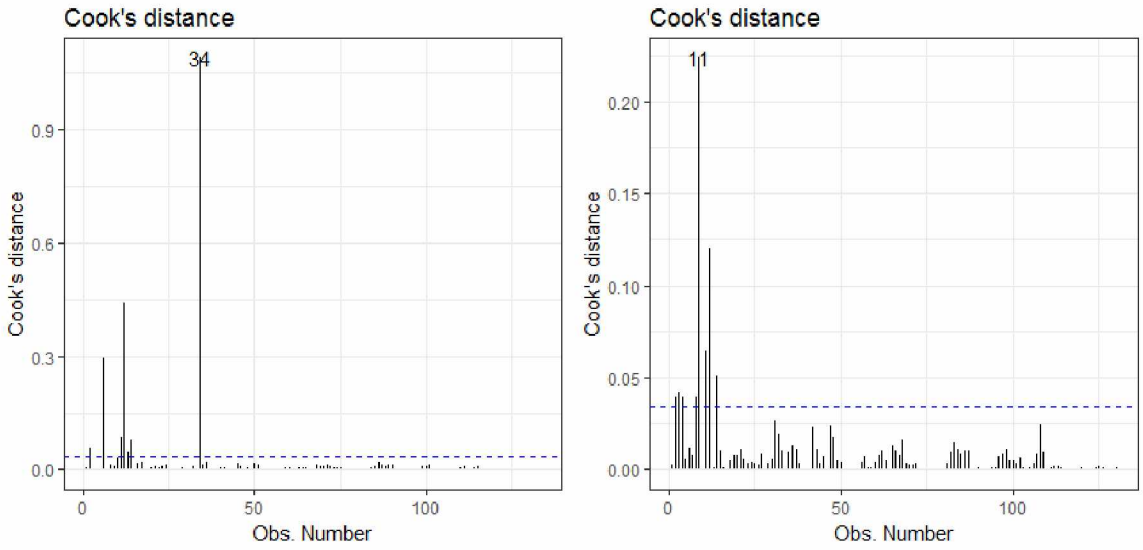

Figure 4-16: Cook's distance plot for five factor model of Proppant Type-1 (Left-before deletion, Right-after deletion)

The regression diagnostic plot before deletion is displayed in **Figure 4-17.** In the residuals vs. fitted plot (upper left), 11^{th}, 14^{th}, and 34^{th} number observations are certainly outliers, but overall there is no distinctive pattern created by the data points. All the data points are almost equally spread around a horizontal line, which indicates that there is no non-linear relationship between the dependent and the independent variables. In the normal Q-Q plot (upper right), most of the data points fall on the line, although there are still a few data points at the top and the bottom including the outliers, which certainly do not fall on that line. The distribution in the normal Q-Q plot (upper right) appears to be normal. In the scale location plot (bottom), there is a random band around the horizontal line, which suggests that the homoscedasticity assumption is met.

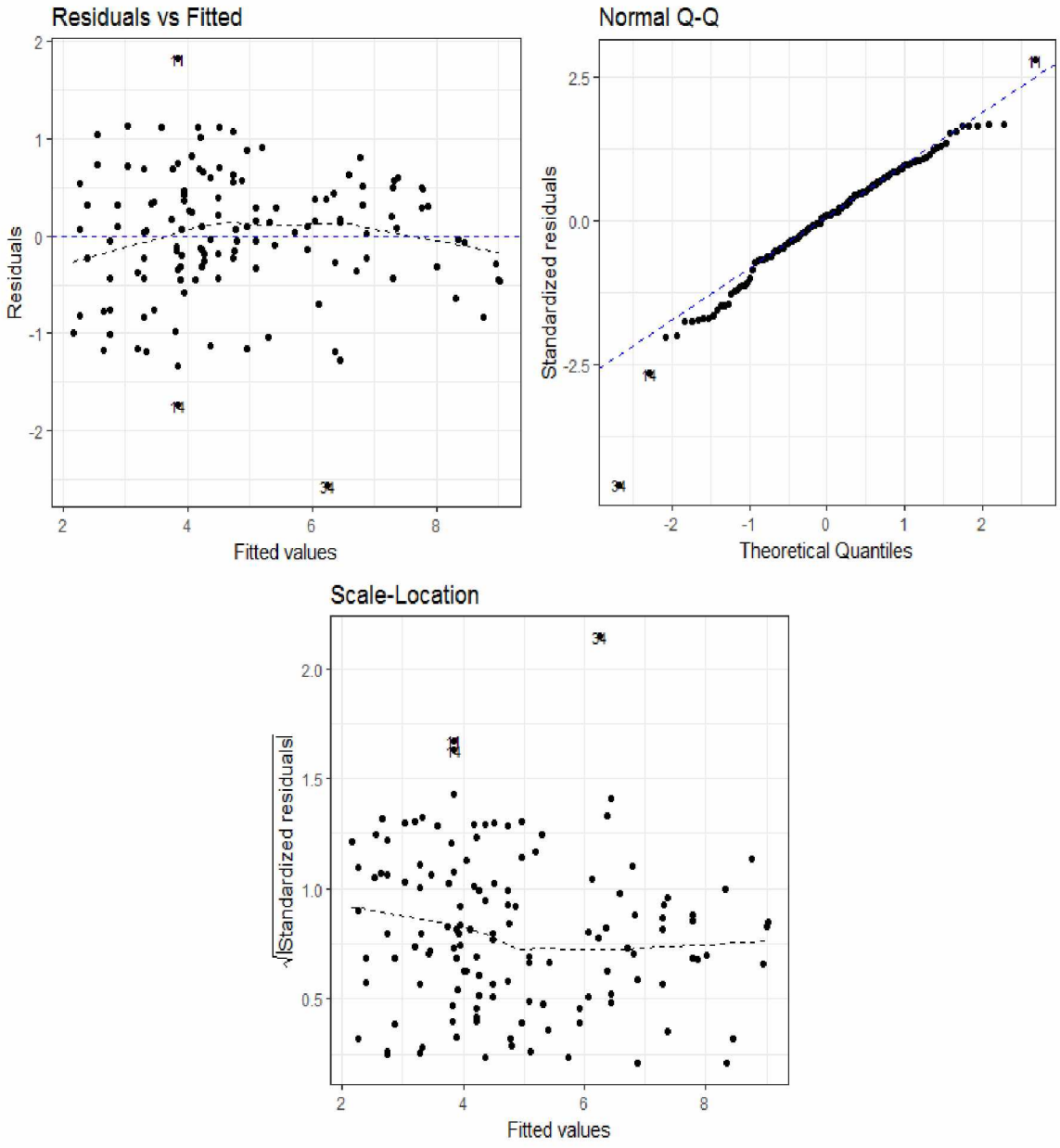

Figure 4-17: Regression diagnostic before deleting influential observations as the validation of model assumptions for five factor model of Proppant Type-1, (Upper left-Linearity, Upper right-Normality, Bottom-Homoscedasticity)

The regression diagnostic plot after deletion is displayed in **Figure 4-18**. There is no major change in the **Figure 4-18** from **Figure 4-17** indicating that the model plausibly meets all the model assumptions.

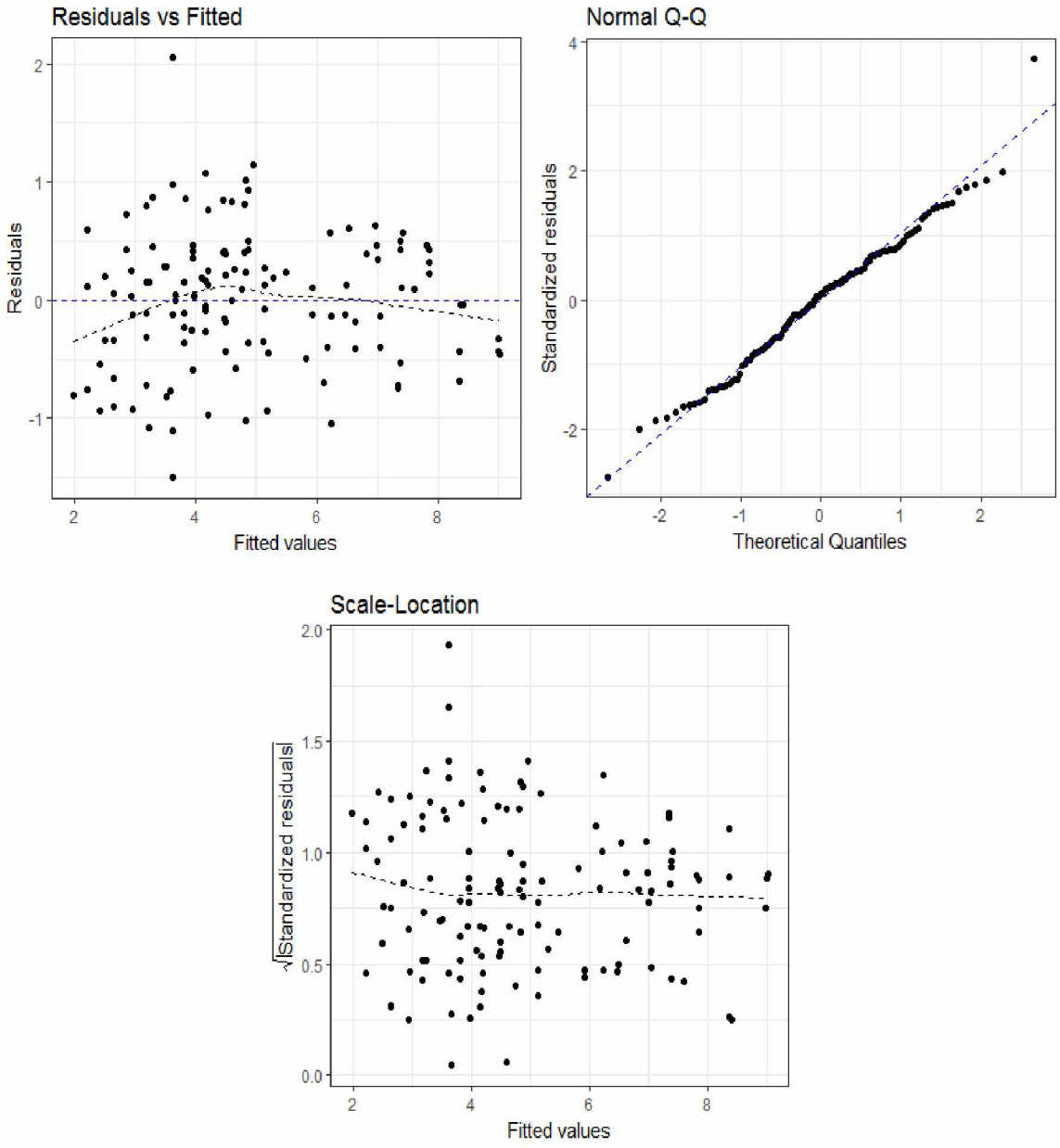

Figure 4-18: Regression diagnostic after deleting influential observations as the validation of model assumptions for five factor model of Proppant Type-1, (Upper left-Linearity, Upper right-Normality, Bottom-Homoscedasticity)

For further analysis, Q-Q plot and distribution of errors plot by studentized residuals are generated to verify whether the assumption of normality is valid or not. Q-Q plot of studentized residuals shown in **Figure 4-19** indicates that most of the data points are inside the 95% confidence envelope. In **Figure 4-20**, the distribution of error plot shows that the normal curve and the kernel

density curve almost overlap each other, so the normality assumption is well satisfactory. In the spread level plot of **Figure 4-21**, the red line is nearly horizontal. The suggested power transformation is 1.07 for this model. Thus, the regression diagnostic suggests that the model assumptions are well satisfied.

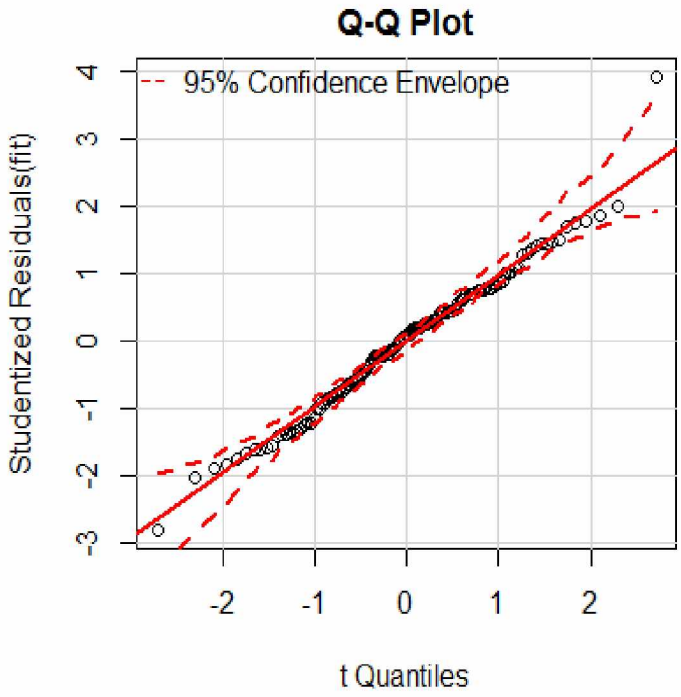

Figure 4-19: Q-Q plot of studentized residuals for five factor model of Proppant Type-1

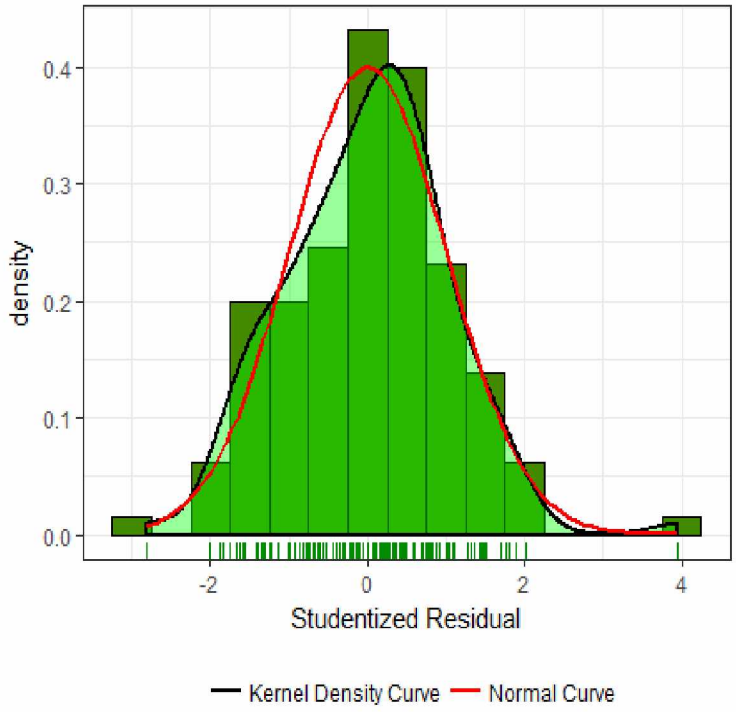

Figure 4-20: Distribution of errors for five factor model of Proppant Type-1

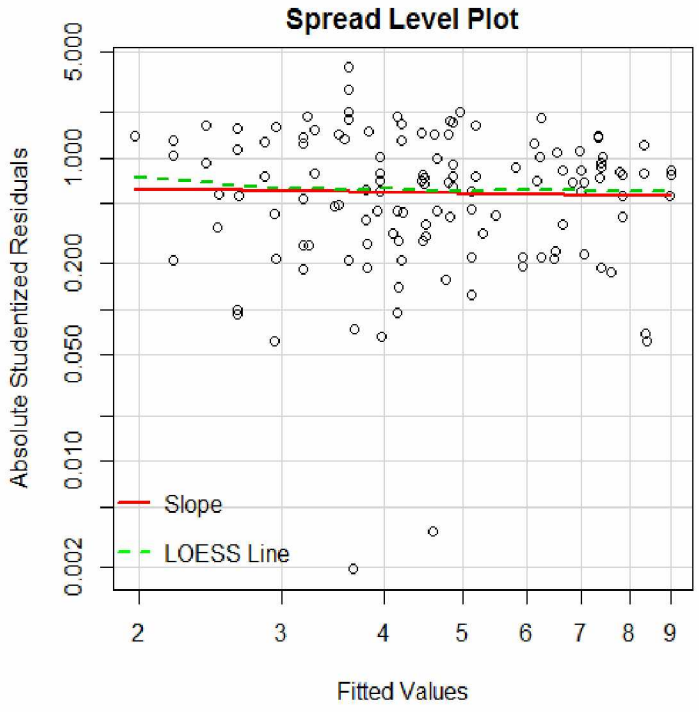

Figure 4-21: Spread level plot for five factor model of Proppant Type-1

After performing 10-fold cross validation, the original R-square and the cross validated R-square are reported as 0.9056 and 0.8826 respectively. The change in R-square is only 0.023, which indicates the model might be capable of providing good predictions on newer datasets. Additionally, the original adjusted R-square and the cross validated adjusted R-square are reported as 0.8977 and 0.8727 respectively. The final model is given below in the **Equation 4.4** where temperature, polymer concentration, and the polynomial term of temperature have p-values higher than 0.05.

$$Fc = \exp[6.93165 + (1.26923 * P_{ms}) - (0.90890 * \sigma_c) + (3.44982 * C_p) - (0.10642 * T) - (0.08941 * C_{pl}) + (0.16575 * {\sigma_c}^2) - (1.90421 * {C_p}^2) + (0.05304 * T^2) - (0.40768 * {C_{pl}}^2)] \quad (4.4)$$

Finally, it is reasonable to state that between all the models developed for Proppant Type-1, the five factor model is the most adequate in terms of model assumptions, model quality, and future predictions. However, the detailed comparison will be discussed in section 4.1.5.

4.1.2 Proppant Type-2

4.1.2.1 Three Factor Model

The graphical representations of the analysis for rest of the fixed models are provided in **Appendix A**. For Proppant type-2, data of mesh size, closure stress, and temperature were available, so only one model is developed using around 140 data points. The initial model has an R-square and an adjusted R-square of 0.7117 and 0.6984 respectively. The optimal model is displayed in **Figure A-8**. All the parameters are included in the best fitting model where R-square is the criterion. **Figure A-9** shows the Cook's distance plot before deletion. Here, deleting the 132th observation does not improve the model, so no observation was removed from the primary dataset.

The basic regression diagnostic is only performed once with the optimal model, which is shown in **Figure A-10**. In the residuals vs. fitted plot (upper right), different curves are formed, indicating a strong violation of linearity assumption. The normal Q-Q plot (upper right) suggests that the distribution is left skewed and it evidently shows that few data points at the bottom do not fall on

the line. The scale-location plot (bottom) shows a nearly parabolic shape, indicating a clear violation of the homoscedasticity assumption. **Figure A-11** and **Figure A-12** display the Q-Q plot and distribution of errors by studentized residuals respectively. The first one demonstrates that a good number of data points are outside the 95% confidence envelope and the latter one suggests a left skewed histogram or negatively skewed histogram. The spread level plot in **Figure A-13** validates the conclusion made from the scale-location plot. The suggested power transformation for this model is 8.02, which is exceedingly high.

From 10-fold cross validation, the original R-square and the cross validated R-square are reported as 0.7117 and 0.5355 respectively. Additionally, the original adjusted R-square and the cross validated R-square are 0.6961 and 0.5103 respectively. The model is proposed by the following **Equation 4.5** where all the parameters are significant.

$$Fc = exp[8.09765 - (0.19166 * P_{ms}) - (0.34484 * \sigma_c) - (0.37392 * T) - (0.31527 * P_{ms} * \sigma_c) + (0.10521 * M * T) - (0.47903 * \sigma_c * T)] \quad (4.5)$$

The effect of mesh size in **Equation 4.5** is reported as negative. Therefore, in addition to all the violations of model assumptions and failing to provide better prediction, the model also shows discrepancy from the established literature.

4.1.3 Proppant Type-3

4.1.3.1 Three Factor Model

There are 675 data points used to develop the three factor model of Proppant Type-3. The first linear regression reports a negative main effect of mesh size. So, polynomial regression of 2nd order is employed to develop the model which has an R-square of 0.3436 and an adjusted R-square of 0.3377. The results of all subsets regression are displayed in **Figure A-14** where R-square is the selection criterion. To reduce the number of parameters in the optimal model, 2nd best fitting model is chosen from the plot of **Figure A-14** where polynomial term of closure stress is left out. The Cook's distance plot is shown in **Figure A-15**. Similar to the three factor model of Proppant Type-2, no observation is deleted. The final model has an R-square and an adjusted R-square of 0.3436 and 0.3387 respectively.

Figure A-16 contains initial regression diagnostics. In the residuals vs. fitted plot (upper left), the residual cloud forms a funnel shape with shorter end toward larger fitted values. In the normal Q-Q plot (upper right), an inverted S-curve is formed around the line, which suggests a distribution with light tails. The square root of standardized residuals create a nearly funnel shape with a downward bend in the scale-location plot (bottom). Furthermore, normality is assessed by **Figure A-17** and **Figure A-18**. **Figure A-17** shows the Q-Q plot by studentized residuals where a large portion of the data points is outside of the 95% confidence envelope. Distribution of the studentized residuals in **Figure A-18** shows a left-skewed histogram suggesting deviation from the normal curve. The spread level plot in **Figure A-19** displays a sharp downward non-horizontal red line implying violation from constant variance assumption. The suggested power transformation is 4.86 for this case. Thus, the model shows strong violation from all the model assumptions.

The original R-square and the cross validated R-square are 0.3436 and 0.3271 respectively. And the original adjusted R-square and the cross validated adjusted R-square are 0.3377 and 0.3211 respectively. The final model is proposed by the following **Equation 4.6** where all the parameters are significant.

$$Fc = exp[8.64340 + (0.41235 * P_{ms}) - (0.28032 * \sigma_c) - (0.31361 * T) - (0.21124 * P_{ms}{}^2) - (0.07990 * T^2)] \quad (4.6)$$

4.1.3.2 Four Factor Model

The model is developed using around 300 data points. Polynomial regression of 2nd order is employed. The initial model has an R-square of 0.6954 and an adjusted R-square of 0.687. The optimal model is selected using adjusted R-square as the criterion. The optimal model is shown in **Figure A-20** where the polynomial terms of closure stress and temperature are excluded. **Figure A-21** shows the Cook's distance plot before deletion (left) and after deletion (right). Altogether 8 observations are removed to reach the final model. This model has an R-square and an adjusted R-square of 0.8033 and 0.7991 respectively. Deleting 8 observations has decreased the mean squared error (MSE) from 0.33 to 0.18. The regression diagnostics before deletion and after deletion are displayed in **Figure A-22** and **Figure A-23** respectively. The residual cloud in the residuals vs.

fitted plot (upper left) is more scattered in **Figure A-23** than it is in **Figure A-22**, but in both plots, the residual cloud has a front nearly ellipsoidal shape toward larger fitted values. The normal Q-Q plots (upper right) of both of these figures have an inverted S-curve formed around the line with heavy tails. A downward residual cloud is formed in both the scale location plots (bottom) but the cloud is more scattered for the 2nd case. Normality and homoscedasticity are further assessed by Q-Q plot, a distribution of error plot and a spread level plot using studentized residuals, which are displayed in **Figure A-24**, **Figure A-25,** and **Figure A-26** respectively. However, all of these imply strong violation from normality and homoscedasticity assumptions. The suggested power transformation is reported as 5.52.

Despite the strong violation from the model assumptions, this model has a cross validated R-square of 0.7850 where the original R-square is 0.8032. The cross validated adjusted R-square and the original adjusted R-square are 0.7805 and 0.7984 respectively. However, satisfactory results from cross validation cannot evidently prove the predictive performance of a regression model until the model satisfies all the model assumptions. **Equation 4.7** displays the model where the main effect of mesh size has p-value higher than 0.05.

$$Fc = exp[8.57212 + (0.06329 * P_{ms}) - (0.53543 * \sigma_c) + (1.53781 * C_p) - (0.09027 * T) - (0.37050 * {P_{ms}}^2) - (0.44576 * {C_p}^2)] \quad (4.7)$$

4.1.3.3 Five Factor Model

Around 90 data points are used to develop the initial linear regression model. Polynomial regression of 2nd order is employed to flip the incorrect sign of the main effect of mesh size. This polynomial model has an R-square of 0.6406 and an adjusted R-square of 0.5946. All subsets regression is conducted using R-square as the selection criterion. The optimal model is shown in **Figure A-27** where the best fitting model does not include the main effect of temperature. However, temperature is included because the polynomial term of temperature is present in the best fitting model. Only the polynomial term of closure stress is left out from the optimal model. Cook's distance plot is displayed in **Figure A-28**. No influential observation was needed to remove

from the original dataset. The final optimal model has an R-square and an adjusted R-square of 0.6395 and 0.5984 respectively. The mean squared error (MSE) of the model is 0.79.

The initial regression diagnostic is provided in **Figure A-29**. In the residuals vs. fitted plot (upper left), all the residuals are distributed well around the horizontal line. It is a clear indication of meeting the linearity assumption. The normal Q-Q plot (upper right) shows that most of the data points do fall on the line and the rest fall near that line. Therefore, the data points are normally distributed. The scale location plot (bottom) shows a well scattered residual cloud. **Figure A-30** and **Figure A-31** show the Q-Q plot and distribution of errors by studentized residuals. From **Figure A-30**, it is apparent that almost all the data points are inside the 95% confidence envelope. **Figure A-31** shows nearly overlapping normal and kernel density curve. Therefore, this model satisfies the normality assumption as well. The red non-horizontal line of the spread level plot in **Figure A-32** suggests a slight violation of constant variance assumption. The suggested power transformation is 1.81 for this model.

The results of 10-fold cross validation show an original R-square of 0.6395 and cross validated R-square of 0.5096. The original adjusted R-square and the cross validated adjusted R-square are 0.5932 and 0.4467 respectively. The model is proposed by the **Equation 4.8** where the main effects of temperature and polymer concentration as well as of polynomial term of polymer concentration have p-values higher than 0.05.

$$Fc = exp[8.22287 + (0.78182 * P_{ms}) - (0.86664 * \sigma_c) + (1.70340 * C_p) - (0.05282 * T) - (0.03940 * C_{pl} - (0.87666 * {P_{ms}}^2) - (0.61180 * {C_p}^2) + (0.33036 * T^2) + (0.07949 * {C_{pl}}^2)] \quad (4.8)$$

From all the above discussions, it can be deduced precisely that the five factor model has met almost all the model assumptions compared to the other two models. It should also be noted that though the four factor model has higher R-square and adjusted R-square than the five factor model, but still the five factor model is more adequate than the four factor model.

4.1.4 All Proppant Type

4.1.4.1 Six Factor Model

This model is developed including types of proppant as a categorical variable along with the other five continuous variables of Proppant Type-1 and Proppant Type-3. Proppant Type-1 is categorized as "1" and Proppant Type-3 as "2". Around 240 data points were available for the initial linear model. However, polynomial regression of 2^{nd} order is employed to flip the incorrect sign of the main effect of mesh size. The very first polynomial model has an R-square and an adjusted R-square of 0.8394 and 0.8316 respectively. The regression coefficient of polynomial term of types of proppant is reported as "NA." Therefore, this parameter is excluded from the model. The optimal model is displayed in **Figure A-33** where R-square is the selection criterion. The optimal model does not include the main effect of types of proppant and polynomial terms of mesh size and polymer concentration. However, the main effect of types of proppant is not excluded from the final optimal model. **Figure A-34** shows the Cook's distance plot before (left) and after (right) deletion. Nine influential observations are deleted starting from 204^{th} number observation. After deleting these influential observations the mean squared error (MSE) is reduced from 0.69 to 0.52. The R-square and adjusted R-square of this final model are 0.8679 and 0.8625 respectively.

The regression diagnostic plots are shown in **Figure A-35** and **Figure A-36** for cases of before deletion and after deletion respectively. The residual cloud has a very mild parabolic shape formed in the residual vs. fitted plot (upper left) of **Figure A-35**. On the other hand, the same plot of **Figure A-36** shows that the residual cloud becomes more scattered and the parabolic shape has nearly disappeared. The normality assumption is met more appropriately in the Q-Q plot (upper right) of **Figure A-36** than of **Figure A-35**. In both of the scale location plots (bottom) of **Figure A-35** and **Figure A-36**, the homoscedasticity assumption seems to be satisfied. The Q-Q plot (**Figure A-37**) and distribution of errors plot (**Figure A-38**) by studentized residuals validate the claim of normality assumption being met. A nearly horizontal red line in the spread level plot (**Figure A-39**) validates the early statement of satisfying homoscedasticity assumption. The suggested power transformation for this model is 0.88.

The cross validated R-square and the original R-square are reported as 0.8596 and 0.8679 respectively. The change in R-square is only 0.008, so the model might perform well with new

datasets. Additionally, the cross validated adjusted R-square and the original adjusted R-square are reported as 0.8531 and 0.8619 respectively. The final model is proposed by **Equation 4.9**. In this equation, only types of proppant has a p-value greater than 0.05.

$$Fc = exp[5.29372 + (1.50163 * P_{ms}) - (0.92900 * \sigma_c) + (1.47756 * C_p) - (0.49934 * T) - (0.57253 * C_{pl} + (0.26464 * t_P) + (0.22944 * {\sigma_c}^2) - \left(0.69801 * {C_p}^2\right) + (0.67987 * T^2)] \quad (4.9)$$

4.1.5 Model Validation

4.1.5.1 Proppant Type-1

All the models developed for Proppant Type-1 are validated by the 30% validation data which was separated at the beginning of this research. More than 100 data points are used to validate the 2nd four factor model. On the other hand, 43 data points are used to validate the rest of the models developed for Proppant Type-1. **Figure 4-22** is the graphical representation of predictions on validation data with 95% prediction interval for these three models. **Equation 4.1** is used to estimate the prediction interval (PI) where $\hat{Y}_i$ is the predicted natural log of fracture conductivity, $t_{crit}$is the critical value of t distribution, MSE is the mean squared error of residuals, X is the $(p+1) \times 1$ column vector of variables, and X_0 is the $(p+1) \times 1$ column vector of new values of variables. Here, p is the number of variables.

$$PI = \hat{Y}_i \pm t_{crit}\sqrt{MSE\left(1 + {X_0}^T (X^T X)^{-1} X_0\right)} \quad (4.10)$$

For all the models, most of the actual conductivity data fall inside the prediction interval. Compared to the others, however, the bottom plot (predictions from five factor model) shows more accurate predictions. In the bottom plot of **Figure 4-22**, the actual conductivities are closer to the prediction base line and only one or two of data points are outside the prediction interval. On the

other hand, the actual conductivity is more scattered from prediction baseline for the other two plots at the top. The prediction interval envelope is also tapered for the five factor model than the other two. The mean squared prediction errors (MSPE) for three, four, and five factor models are 3.42, 2.91, and 0.48 respectively. Precisely, the five factor model provides more accurate predictions than the other two models.

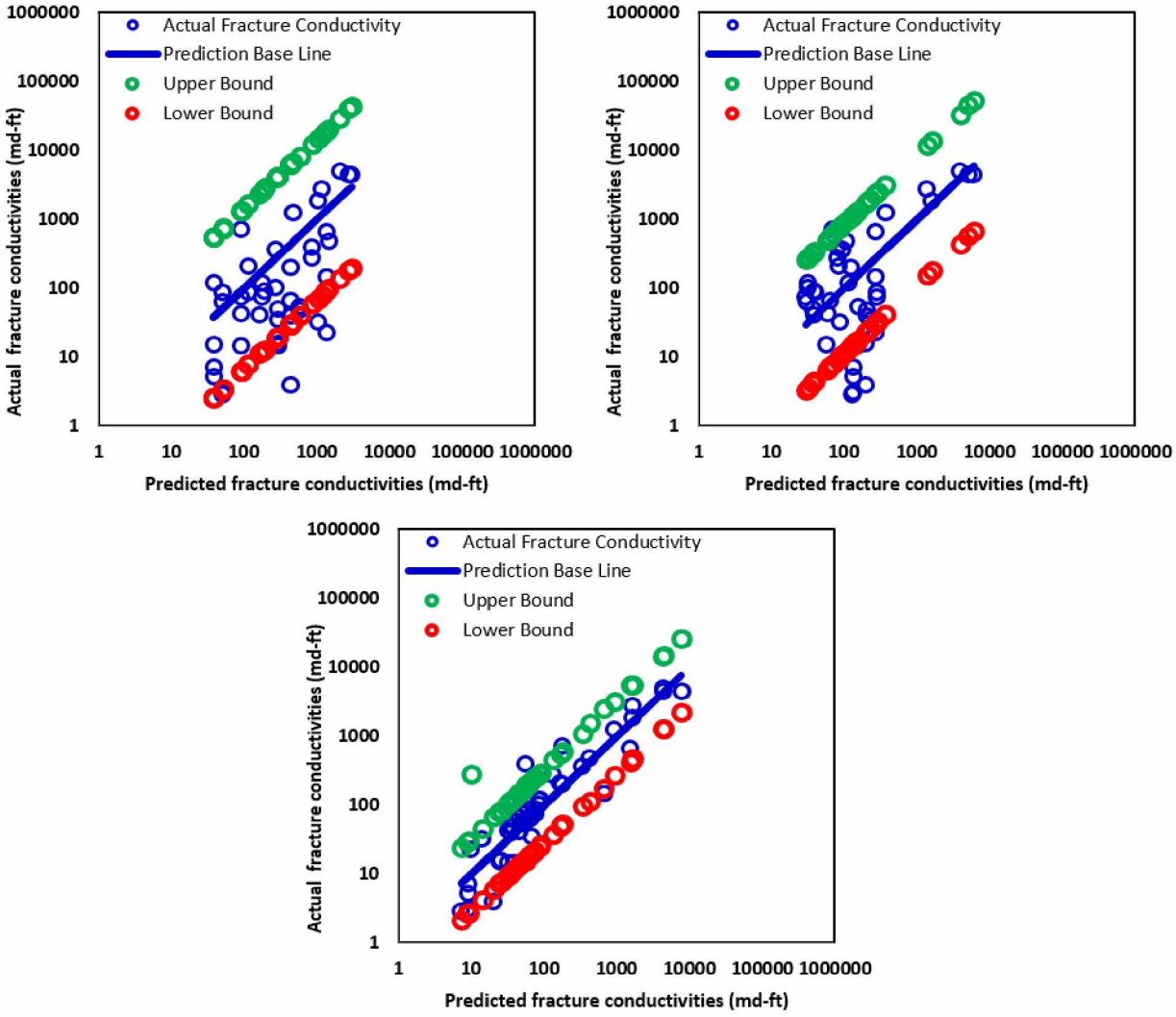

Figure 4-22: Graphical representation of predictions for the models of Proppant Type-1 on 30% validation data with 95% prediction interval (Upper left-three Factors, Upper right-1st four factors, Bottom-five factors)

The graphical representation of prediction for the other four factor model is shown in **Figure 4-23**. Actual conductivities are reasonably inside the confidence envelope. The mean squared prediction error (MSPE) for this model is 1.44.

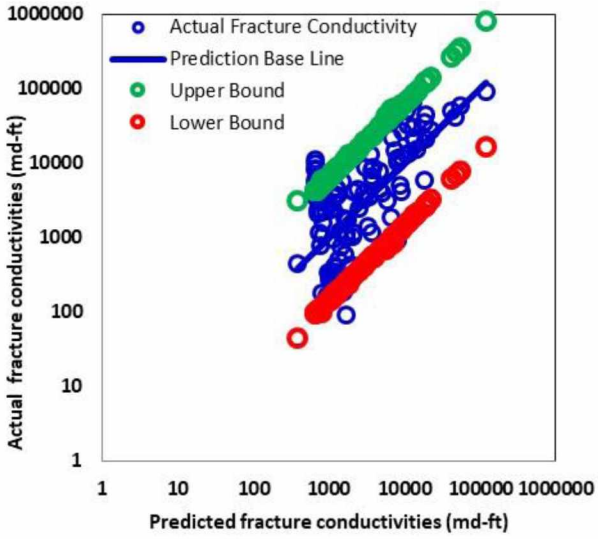

Figure 4-23: : Graphical representation of predictions for the 2nd four factor model of Proppant Type-1 on 30% validation data with 95% prediction interval

4.1.5.2 Proppant Type-2

The graphical representation of predictions for the three factor model of Proppant Type-2 is displayed in **Figure 4-24** where the axis scales are different from **Figure 4-23**. A different axis scale is used to have a clearer view as actual and predicted conductivities are between the ranges of 1000-10000 md-ft. Around 55 data points are used to validate the model. From **Figure 4-24**, it seems reasonable to state that the model predicts showing minimal errors with the validation data. The mean squared prediction error (MSPE) of predictions is only 0.08. From the results of model validation, the model seems very precise on predictions. However, the regression diagnostics performed for this model suggest that it is inadequate in terms of model assumptions. In addition, the model is developed using only three factors as no data was accessible with any additional factors. Altogether, the model is unfit for use. Therefore, the model is not considered for any further analysis.

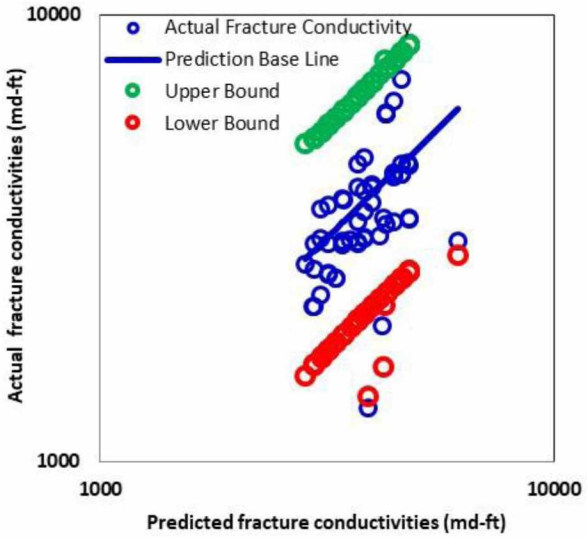

Figure 4-24: Graphical representation of predictions for the three factor model of Proppant Type-2 on 30% validation data with 95% prediction interval

4.1.5.3 Proppant Type-3

There are 43 data points used to validate the rest of the models developed for Proppant Type-3. **Figure 4-25** is the graphical representation of predictions on validation data with a 95% prediction interval for these three models. The upper left and upper right plots of **Figure 4-25** show predictions for three and four factor model respectively. Three or more data points are outside the prediction interval in these two plots. In the upper left plot, actual conductivities are widely scattered around the prediction baseline. The upper right plot has a more tapered prediction interval compared to the others. On the other hand, the plot at the bottom of **Figure 4-25** shows more accuracy in predictions. Only a single data point is outside the prediction envelope in this plot. Actual conductivities are well around the prediction baseline. The mean squared prediction errors (MSPE) for three, four, and five factor models are 1.61, 0.65, and 0.80 respectively. Evidently, the four and five factor model provide better predictions than the three factors model.

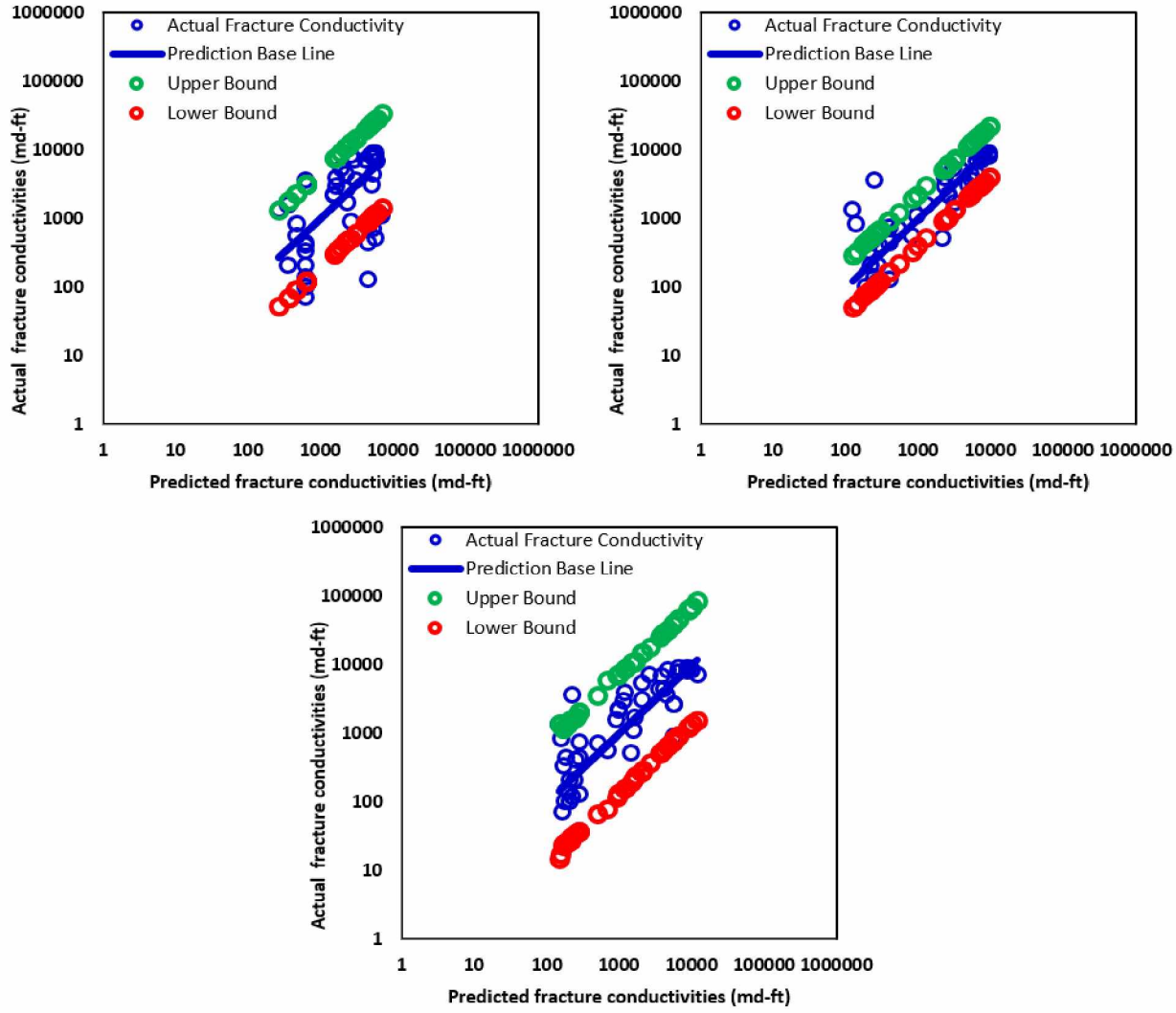

Figure 4-25: Graphical representation of predictions for the models of Proppant Type-3 on 30% validation data with 95% prediction interval (Upper left-three Factors, Upper right-1st four factors, Bottom-five factors)

4.1.5.4 All Proppant Type

All Proppant Type is the mixed proppant types including both sand and ceramic proppants. **Figure 4-26** represents the prediction on validation data for this model. Around 115 data points are used to validate the model. More than 90% of actual conductivities are inside the 95% prediction envelope and they are scattered closely around the prediction baseline. The mean squared prediction error (MSPE) is 1.10 for this case.

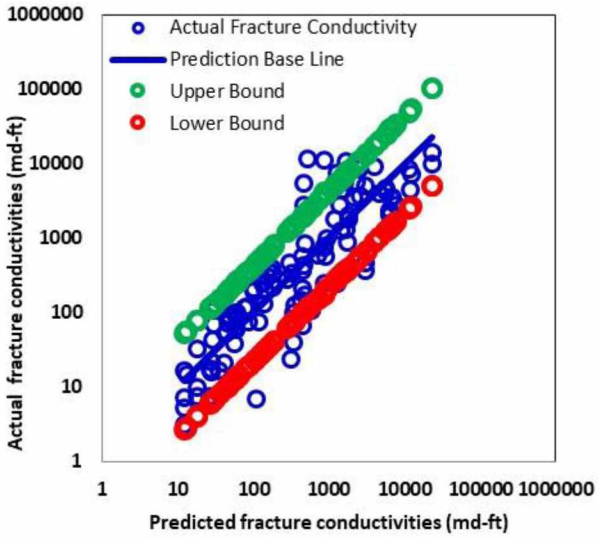

Figure 4-26: Graphical representation of predictions for the six factor model of All Proppant Type on 30% validation data with 95% prediction interval

4.1.6 Model Comparison-Three vs. Four vs. Five Factor Model

4.1.6.1 Proppant Type-1

The performance of all three models of Proppant Type-1 is summarized in **Table 4-1**. The percent decreases in cross validated R-square from original R-square are listed as 38%, 31%, and 2.5% for three, four, and five factor model respectively. Apparently, the five factor model shows better predictions with newer data than the other two models. Moreover, the mean squared error (MSE) and mean squared prediction error (MSPE) are the lowest for the five factor model. Additionally, the model has the value of suggested power transformation closer to 1. In addition, the model validation shown in **Figure 4-22** also suggests that the five factor model is more efficient. The model also satisfies all the model assumptions more appropriately. Hence, these three models can be arranged in order of satisfying model adequacy from low to high: four, three, and five factor model. Precisely, the five factor model is the best amongst these three.

Table 4-1: Comparison of model performances between three, four and five factor models of Proppant Type-1

Performance Assessing Measure		Three Factor Model	Four Factor Model	Five Factor Model
K-fold cross validation	Original R-square	0.5206	0.7442	0.9056
	Cross validated R-square	0.3219	0.5106	0.8826
	Decrease in R-square (%)	38%	31%	2.5%
MSE		1.83	1.18	0.33
MSPE		3.42	2.91	0.48
Suggested Power Transformation		1.895	2.09	1.07
Satisfaction levels of Model Assumptions	Linearity	Moderate	Low	High
	Normality	High	Moderate	High
	Homoscedasticity	Moderate	Low	High
Overall Model Adequacy		Fair	Poor	Good

4.1.6.2 Proppant Type-3

The comparison of model performances between all the models of Proppant Type-3 is summarized in **Table 4-2**. The percent decrease from original R-square to cross validated R-square is the highest for the five factor model and the lowest for the four factor model. The mean squared error (MSE) and mean squared prediction error (MSPE) are also lowest for the four factor model and highest for the three factor model. Thus, the five factor model shows intermediate performance. Both the three and four factor model, however, rigorously fail to satisfy all the model assumptions. Both these models have poor model adequacy. The suggested power transformation for these models is extremely high from 1 indicating strong violation of constant variance assumption. On

other hand, the value of power transformation for five factor model is listed as 1.81 which is the closest to 1. Therefore, even though the five factor model is intermediary of all the models in terms of R-square and mean squared error, the model is still the most satisfactory one as it greatly satisfies the model assumptions. Therefore, the five factor model is the best amongst these three.

Table 4-2: Comparison of model performances between three, four and five factor models of Proppant Type-3

Performance Assessing Measure		Three Factor Model	Four Factor Model	Five Factor Model
K-fold cross validation	Original R-square	0.3436	0.8032	0.6395
	Cross validated R-square	0.3271	0.7850	0.5096
	Decrease in R-square (%)	4.8%	2.3%	20%
MSE		0.64	0.18	0.79
MSPE		1.6	0.65	0.80
Suggested Power Transformation		4.86	5.52	1.81
Satisfaction levels of Model Assumptions	Linearity	Low	Low	High
	Normality	Low	Low	High
	Homoscedasticity	Low	Low	High
Overall Model Adequacy		Poor	Poor	Good

4.1.7 Relative Weights of the Predictors

Relative weights of the main variables (excluding other polynomial terms) for Proppant Type 1 and Proppant Type-3 are shown by bar charts in **Figure 4-27** and **Figure 4-28** respectively. The results for both proppants suggest that these two variables consistently contribute the most to the R-square. **Figure 4-27** demonstrates that for Proppant Type-1, mesh size and proppant concentration contribute around 22% and 17% to the R-square respectively, which are the top two contributors. On other hand, for Proppant Type-3, mesh size and proppant concentration are also ranked as the highest contributors with contribution around 28% and 18% respectively. Therefore, it is sensible to claim that mesh size and proppant concentration could be relatively more important than the other variables. For both the models, p-values of temperature and polymer concentration are above 0.05, which also suggests that these two are the less significant variables.

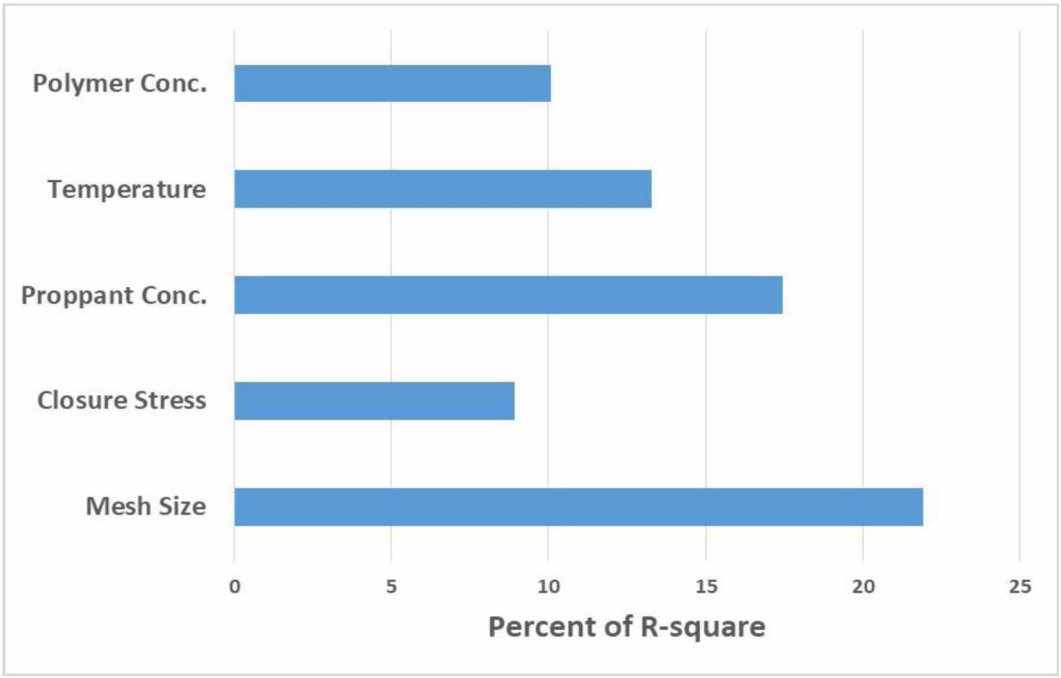

Figure 4-27: Relative weights of the main predictors of the five factor model of Proppant Type-1

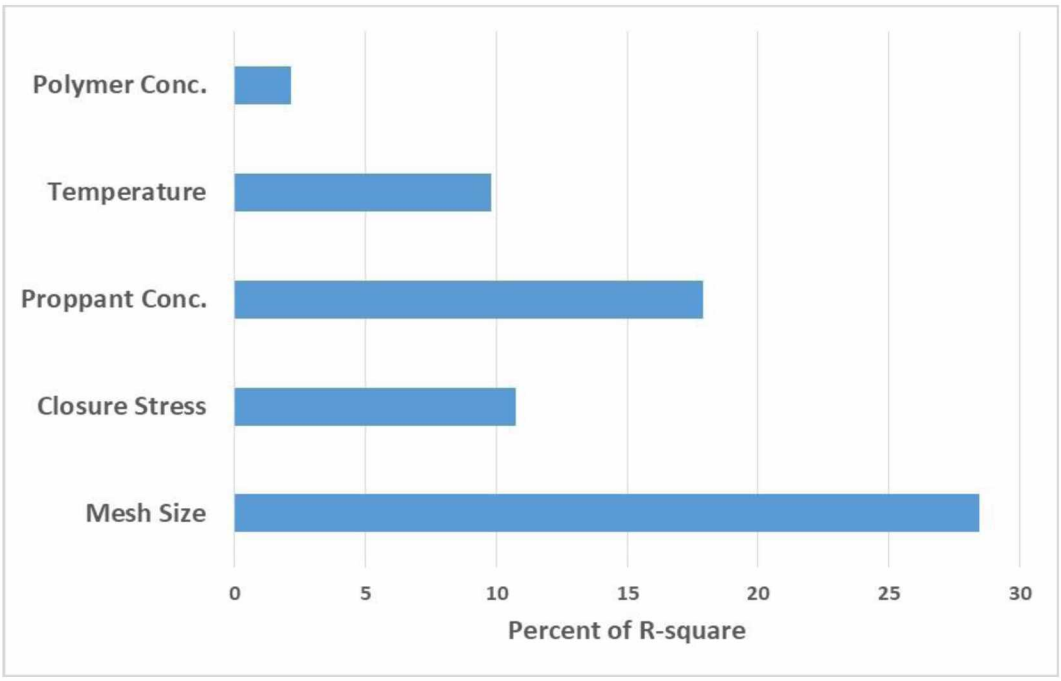

Figure 4-28: Relative weights of the main predictors of the five factor model of Proppant Type-3

4.2 Mixed Effect Model

The aim of using this statistical approach is to improve the five factors fixed effect models developed for Proppant Type-1 and Proppant Type-3. In a mixed effect model, along with all the fixed effects, the results will also display random effects for each group or level of a categorical variable. It is assumed that the random effects are normally distributed. Mixed effect modeling will be carried out with the same predictors that are finally chosen for fixed effect models. However, the main intention of using a mixed effect model is to develop a model which can be used for further prediction and has a lower mean squared error (MSE) than the fixed effect models.

4.2.1 Proppant Type-1

For Proppant Type-1, the entire dataset has five different sources. Each source is labeled by a categorical variable as a level or group. Therefore, there are five groups and it is expected that five random effects will be estimated by the mixed effect model. **Figure 4-29** displays the estimation of random effects for each group. In the x-axis, BLUP is defined as the best linear unbiased prediction, which is the estimation of random effects. From the magnitude of the random effects, it is very obvious that the random effects have very little effect on the intercept. Therefore, a very

little change is expected in the mixed model compared to the fixed model. And for both the cases of Proppant Type-1, there is minimal difference in the relationship between fracture conductivity and its affecting factors among the different source.

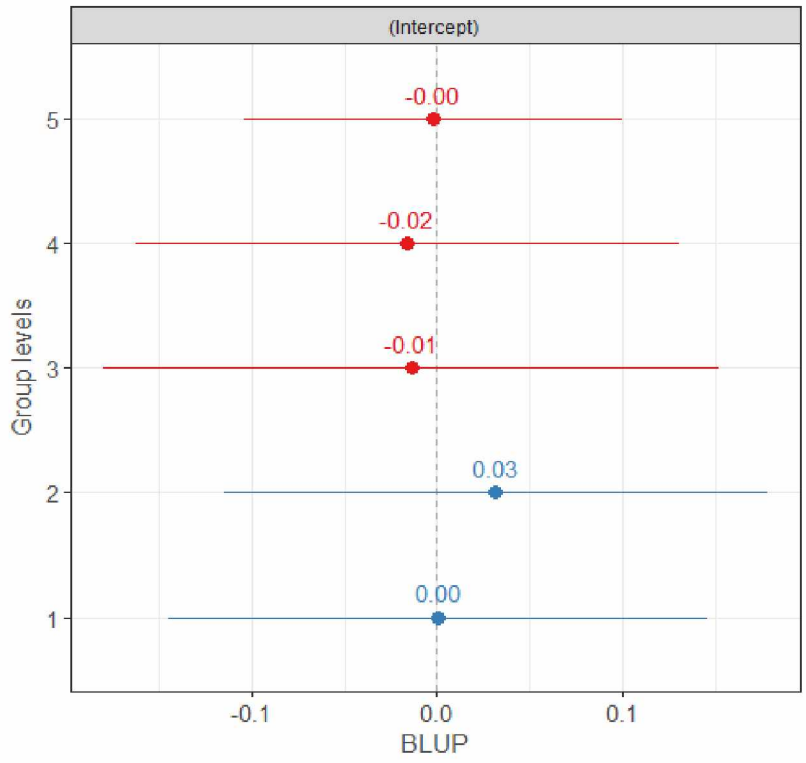

Figure 4-29: Estimations of random effects for mixed model of Proppant Type-1

Figure 4-30 displays the distribution of random effects. The red envelope indicates the 95% confidence interval. **Figure 4-31** shows the marginal effects of the model predictors. The figure suggests that temperature has very minimal effects on hydraulic fracture conductivity. On other hand, the other effects are consistent with the documented literature. The marginal effect of proppant concentration might look a little strange as conductivity does not increase with increasing proppant concentration after a certain point. Overall, however, the marginal effect of proppant concentration indicates that the conductivity should increase with increasing proppant concentration. The mixed effect model equation developed here will evidently support this claim.

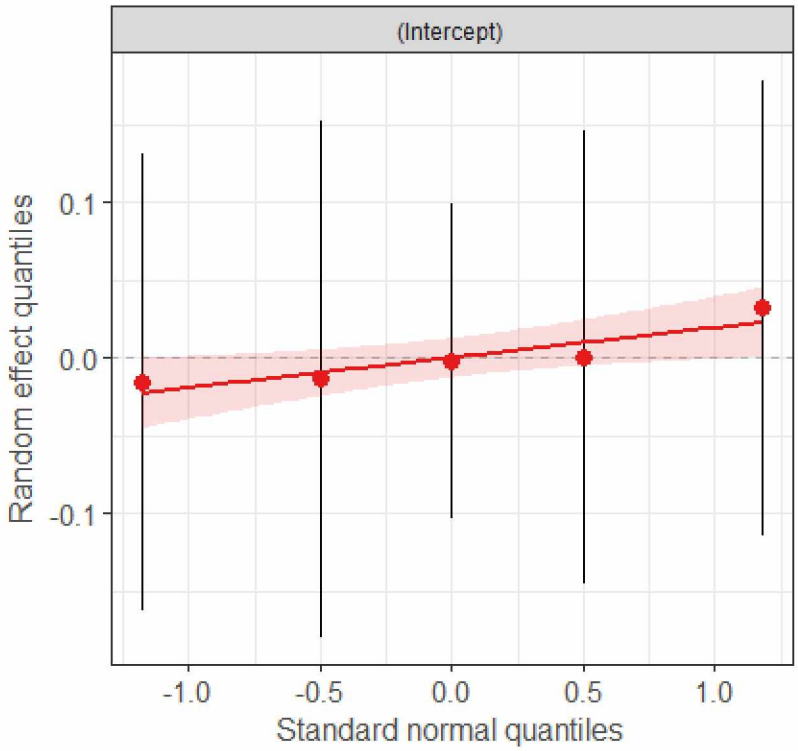

Figure 4-30: Distribution of random effects for mixed model of Proppant Type-1

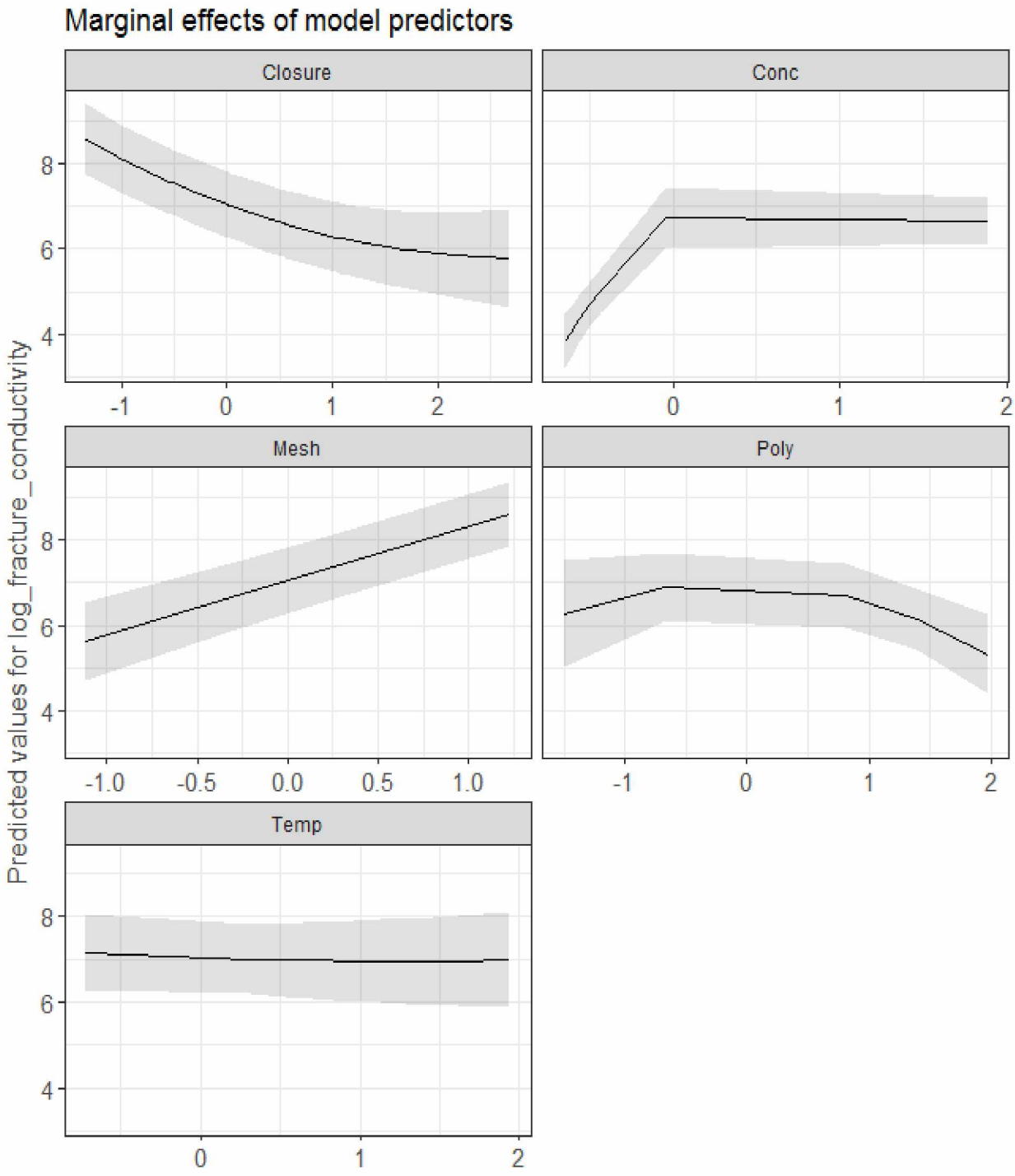

Figure 4-31: Marginal effects of the mixed model predictors of Proppant Type-1

The random effects for the mixed model are presented in **Table 4-3** and the coefficients estimate of intercepts are presented in **Table 4-4**. **Table 4-4** suggests that for each group or level the intercept is different. This is because the intercept includes the random effect.

Table 4-3: Random effects for each level or group of Proppant Type-1

Level or Group	Random Effects
1	0.00027
2	0.0316
3	-0.0139
4	-0.0158
5	-0.0021

Table 4-4: Coefficients estimate of intercepts for each level of Proppant Type-1

Coefficient Estimates	Level or Group				
	1	2	3	4	5
Intercept	6.914524	6.945894	6.900334	6.898403	6.912103

Thus, a grand mean of all the intercepts (**Table 4-4**) is reported. **Equation 4.11** displays the mixed model for Proppant type-1.

$$Fc = exp[6.91425 + (1.27017 * P_{ms}) - (0.90784 * \sigma_c) + (3.45732 * C_p) - (0.13163 * T) - (0.08775 * C_{pl}) + (0.16459 * {\sigma_c}^2) - (1.90147 * {C_p}^2) + (0.05597 * T^2) - (0.39348 * {C_{pl}}^2)] \quad (4.11)$$

The model has a marginal R-square of 0.8971 and a conditional R-square of 0.8992. The marginal R-square is associated with the fixed effect and the conditional R-square is associated with fixed plus the random effects. The mean squared error (MSE) of the model is 0.32.

4.2.2 Proppant Type-3

There are six different sources for the complete dataset of Proppant Type-3, so six random effects are estimated for this mixed model. **Figure 4-32** shows the estimation of random effects and Figure **4-33** shows their distribution. The magnitudes of the random effects are comparatively higher than the previous mixed effect model. Therefore, relatively larger changes are anticipated in this mixed effect model from the fixed effect model of Proppant Type-3. **Figure 4-34** shows the marginal effects of the predictors. The marginal plot indicates that conductivity decreases with increasing closure stress, temperature, and polymer concentration. The magnitudes of changes are comparatively low, however, for temperature and polymer concentration. On other hand, in general mesh size and proppant concentration increase fracture conductivity.

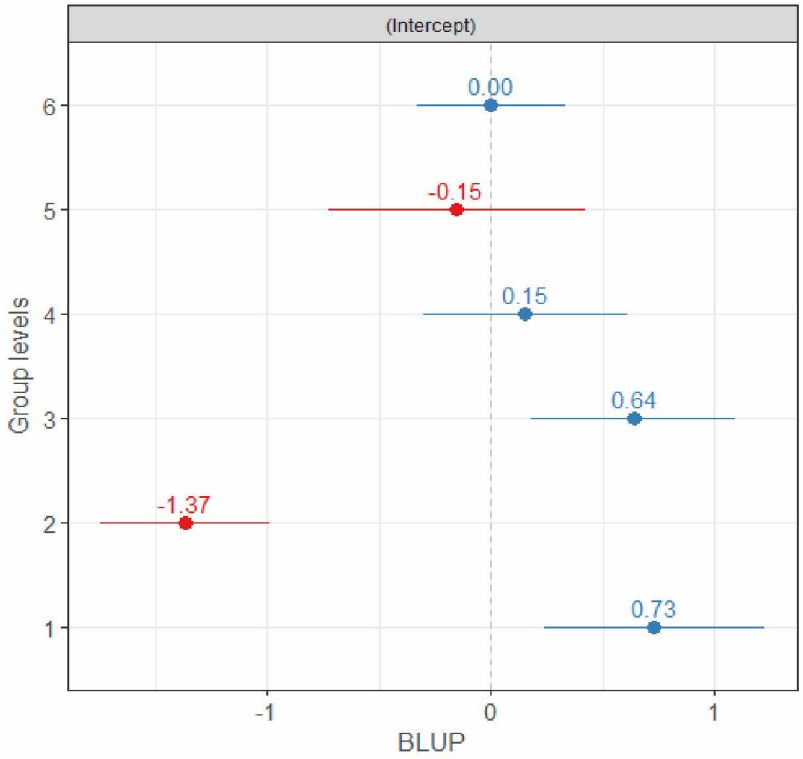

Figure 4-32: Estimations of random effects for mixed model of Proppant Type-3

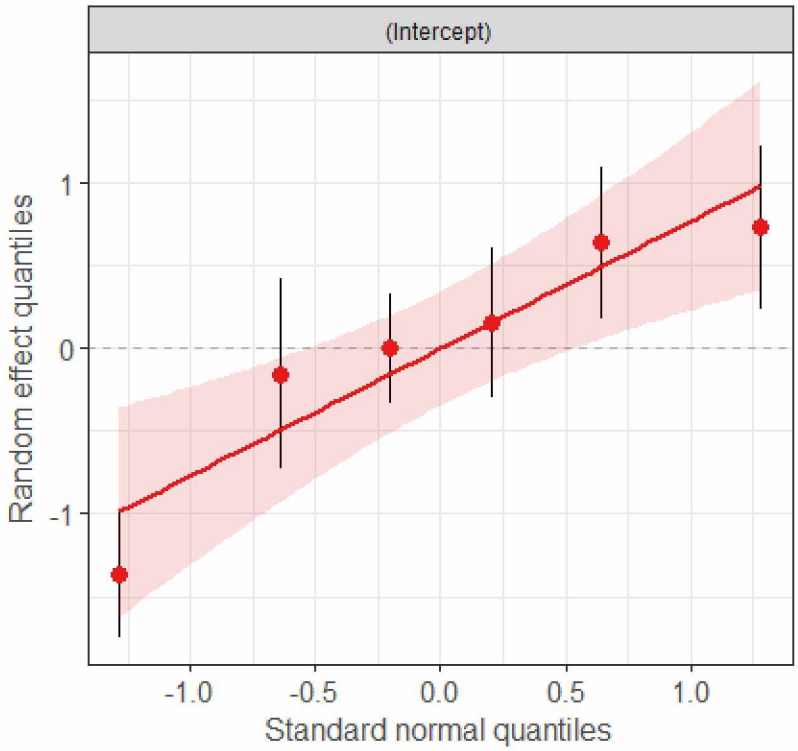

Figure 4-33: Distribution of random effects for mixed model of Proppant Type-3

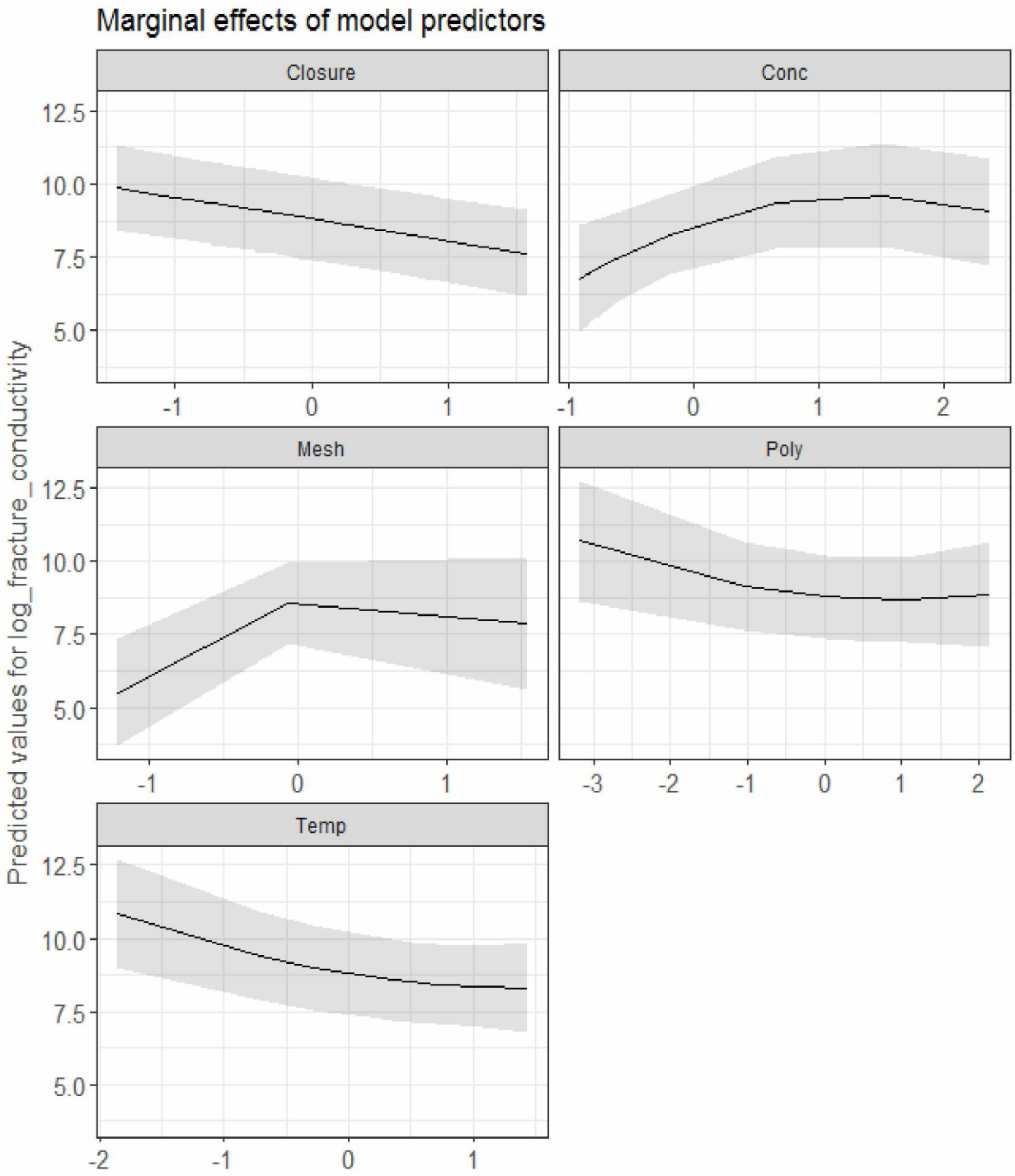

Figure 4-34: Marginal effects of the mixed model predictors of Proppant Type-3

The random effects for the mixed model are presented in **Table 4-5** and the coefficients estimate of intercepts are presented in **Table 4-6**.

Table 4-5: Random effects for each level or group of Proppant Type-3

Level or Group	Random Effects
1	$7.294e^{-01}$
2	$-1.367e^{+00}$
3	$6.375e^{-01}$
4	$1.539e^{-01}$
5	$-1.539e^{-01}$
6	$2.412538e^{-13}$

Table 4-6: Coefficients estimate of intercepts for each level of Proppant Type-3

Coefficient Estimates	Level or Group					
	1	2	3	4	5	6
Intercept	9.289033	7.192617	9.197147	8.713547	8.405651	8.559599

The mixed model with grand mean of all the intercepts (**Table 4-6**) is provided in **Equation 4.12**.

$$Fc = exp[8.55960 + (1.20848 * P_{ms}) - (0.73940 * \sigma_c) + (1.51310 * C_p) - (0.66332 * T) - (0.22407 * C_{pl} - (1.10205 * {P_{ms}}^2) - (0.55294 * {C_p}^2) + (0.23563 * T^2) + (0.11727 * {C_{pl}}^2) \quad (4.12)$$

The model has a marginal R-square and a conditional R-square of 0.4552 and 0.7830. Precisely, it suggests that adding random effects has significant impact on the model. The mean squared error (MSE) for the model is reported as 0.6342.

4.2.3 Model Validation

4.2.3.1 Proppant Type-1

The mixed model is validated with the same validation data that were used to validate the fixed effect model of Proppant Type-1. **Figure 4-35** is the graphical representation of predictions on validation data with 95% prediction interval. **Figure 4-35** seems pretty similar to the bottom plot of **Figure 4-22,** which is the validation plot for five factors fixed effect model of **Proppant Type-1**. The mean squared prediction error (MSPE) of the mixed model is 0.48.

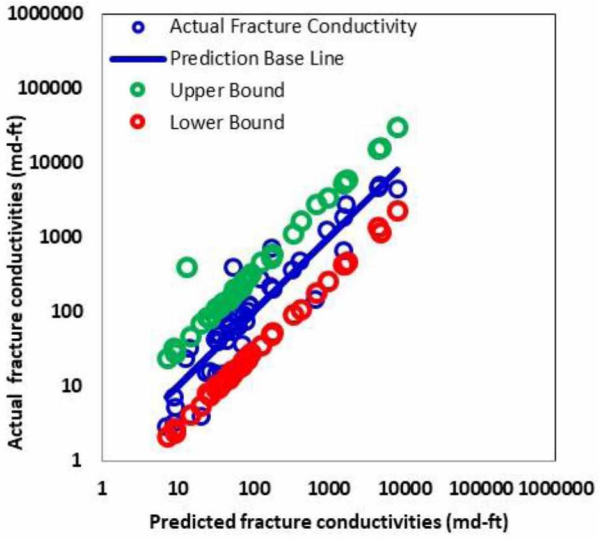

Figure 4-35: Graphical representation of predictions for the five factors mixed model of Proppant Type-1 on 30% validation data with 95% prediction interval

4.2.3.2 Proppant Type-3

The mixed model validation with 30% validation data for Proppant Type-3 is shown in **Figure 4-36**. If **Figure 4-36** is compared with the bottom plot of **Figure 4-25** (model validation of five factors fixed model of Proppant Type-3), then certainly it can be observed that for the mixed model all the actual conductivities are inside the prediction interval. For the fixed model, there is one data point that is outside the prediction interval. However, there has been no significant change from the model validation of fixed effect model. The mean squared prediction error (MSPE) of the mixed model is 0.81.

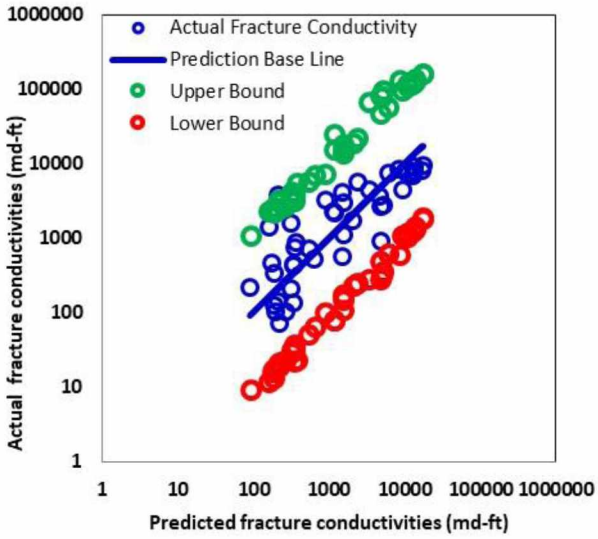

Figure 4-36: Graphical representation of predictions for the five factors mixed model of Proppant Type-3 on 30% validation data with 95% prediction interval

4.2.4 Fixed vs. Mixed Effect Model

Both of the mixed effect models provide almost identical regression diagnostic plots. Thus, mixed effect models for both proppant types could be assumed to satisfy model adequacy. **Table 4-7** presents the comparison between the fixed and the mixed effect models of both Proppant types. The table is mainly based on the mean squared error (MSE) of model and mean squared prediction error (MSPE). For Proppant Type-1, mean squared error (MSE) of the mixed model has decreased by around 3% from the fixed effect model. However, the mean squared prediction errors (MSPE) are same if they are rounded off to two decimal places. On other hand, for Proppant Type-3, mean squared error (MSE) in the mixed model has significantly decreased by around 20% from the fixed effect model. The mean squared prediction error (MSPE), however, is slightly lower in the fixed effect model. All the models are found adequate according to the regression diagnostic plots. Therefore, it can be concluded that the mixed effect models are likely better than the fixed effect models.

Table 4-7: Comparison between fixed and mixed effect models of both proppant types

Performance Assessing Measure	Proppant Type-1		Proppant Type-3	
	Fixed Effect Model	Mixed Effect Model	Fixed Effect Model	Mixed Effect Model
MSE	0.33	0.32	0.79	0.63
MSPE	0.48	0.48	0.80	0.81
Model Adequacy	Good	Good	Good	Good

4.3 Imputation Model

Similar to fixed and mixed effect models, the imputation models will be developed using the same parameters. Imputation will be employed on Proppant Type-1 and Proppant Type-3. At first, the missing data will be imputed by employing two statistical approaches. Then regression analysis will be employed to develop hydraulic fracture conductivity models. For this case, missing values of polymer concentration are imputed using categorical imputation and multiple imputation.

4.3.1 Categorical Imputation

In categorical imputation, the missing data are imputed with categorical variables. Similar to the case of the mixed model, each data source is treated as each group and each group is assigned with a categorical variable that represents a categorical level. A missing predictor value is assigned this new categorical level, distinct from the levels from other sources. It assumes that the value of a categorical variable is consistent throughout a source and its value really is different from the other sources. After imputation, regression analysis is conducted on the imputed dataset. **Appendix A** contains the graphical representations of the regression analysis for the categorical imputation models.

4.3.1.1 Proppant Type-1

After categorical imputation, the method of developing the fracture conductivity model is the same as the fixed effect modeling. The very first regression model developed here has an R-square and an adjusted R-square of 0.5754 and 5688 respectively. Similar to the process of fixed effect

modeling, Cook's distance is employed to find out the influential observations. Only 3 influential observations are deleted from the original dataset by using Cook's distance. After employing Cook's distance, the mean squared error (MSE) has decreased from 2.06 to 2.01. **Figure A-40** shows the Cook's distance plot before and after deletion. The final model has an R-square of 0.5859 and an adjusted R-square of 0.5794.

The regression diagnostic plots before deletion and after deletion are displayed in **Figure A-41** and **Figure A-42** respectively. These plots are quite similar to the regression diagnostic plots of four factor model of Proppant type-1. However, no significant changes occurred after deleting those influential observations. The residual vs. fitted plots (upper left) in both regression diagnostic plots suggest clear violation of linearity assumption. A funnel shape is formed by the residual cloud in the both plots. In the normal Q-Q plot (upper right), a large number of data points do not fall on the line for both cases. The distribution of the data points is ambiguous. And in the both the scale location plots (bottom), the residual cloud is bent downward. **Figure A-43** and **Figure A-44** show further assessments of the normality assumption. **Figure A-43** shows that a large number of the studentized residuals are outside the 95% confidence envelope. And the distribution of errors plot in **Figure A-44** shows a slightly left skewed histogram. The kernel density curve is observed to be deviated from the normal curve, so the model fails to meet the normality assumption. **Figure A-45** shows the spread level plot where a non-horizontal red line is formed by the residual cloud. The power transformation is suggested to be 2.41 indicating deviation from constant variance assumption.

However, from 10-fold cross validation, the cross validated R-square and the original R-square are reported as 0.5674 and 0. 5858 respectively. The cross validated adjusted R-square is 0.5599 and the original adjusted R-square is 0.5786. The model is proposed by **Equation 4.13** where p-values of temperature, polynomial terms of proppant concentration and temperature are above 0.05.

$$Fc = exp[7.36743 + (0.83192 * P_{ms}) - (0.82860 * \sigma_c) + (0.74121 * C_p) - (0.01412 * T) + (0.61678 * C_{pl}) + (0.26926 * {\sigma_c}^2) - (0.05886 * {C_p}^2) - (0.15247 * T^2) - (0.56102 * {C_{pl}}^2)] \quad (4.13)$$

In **Equation 4.13**, the sign before the main effect of polymer concentration is positive. Because the polymer concentration was treated as a category, the interpretation is different for the main effect of polymer concentration. The coefficient of polymer concentration represents the average difference in fracture conductivity between that level of the variable and the reference level.

4.3.1.2 Proppant Type-3

The initial model has an R-square and an adjusted R-square of 0.6644 and 0.6539 respectively. Seven influential observations are removed from the primary dataset by using Cook's distance. The Cook's distance plot is displayed in **Figure A-46**. The mean squared error (MSE) has decreased from 0.36 to 0.17 after employing Cook's distance. The final mode has an R-square of 0.7878 and an adjusted R-square of 0.781.

The regression diagnostic plots before deletion and after deletion are displayed in **Figure A-47** and **Figure A-48** respectively. **Figure A-47** clearly indicates violations of the model assumptions. A nearly funnel shape is formed with ellipsoidal front towards larger fitted values in the residuals vs. fitted plot (upper left). A nearly inverted S-shape is formed around the line in the normal Q-Q plot (upper right). It suggests that the distribution is heavy tailed. The scale location plot (bottom) suggests a downward horizontal bend. There have been no significant changes in **Figure A-48** from **Figure A-47**. Only the residuals became more scattered in the residuals vs. fitted and scale location plots in **Figure A-48**. The Q-Q plot of studentized residuals in **Figure A-49** is same as the normal Q-Q plot (upper right) of **Figure A-48**. The histogram of studentized residuals in **Figure A-50** is left skewed. A sharp downward non-horizontal red line is observed in the spread level plot in **Figure A-51**. The suggested power transformation is 4.34 for this case. Precisely, the model does not meet the model assumptions.

Furthermore, the cross validated R-square and the original R-square are 0.7687 and 0.7877 respectively. Additionally, the cross validated adjusted R-square and the original adjusted R-square are 0.7605 and 0.7802 respectively. The model is proposed by the following **Equation 4.14** where the polynomial term of temperature has a p-value greater than 0.05

$$Fc = exp[8.54424 + (0.30032 * P_{ms}) - (0.53578 * \sigma_c) + (0.81348 * C_p) - (0.06157 * T) - (0.07549 * C_{pl}) - (0.30278 * {P_{ms}}^2) - (0.28091 * {C_p}^2) - (0.03320 * T^2) - (0.11691 * {C_{pl}}^2)] \quad (4.14)$$

4.3.2 Multiple Imputation

According to the section 3.4.2, multiple imputation using chained equation (MICE) is employed for two thousand sample runs. The ultimate interest of the approach is to find out the relationship between hydraulic fracture conductivity and its affecting factors. Therefore, the intent of this analysis is not to obtain the imputed dataset but on estimating the regression coefficients.

4.3.2.1 Proppant Type-1

Figure 4-37 displays the proportion of missing values of polymer concentration. Around 81% of values of polymer concentration are missing from the entire dataset.

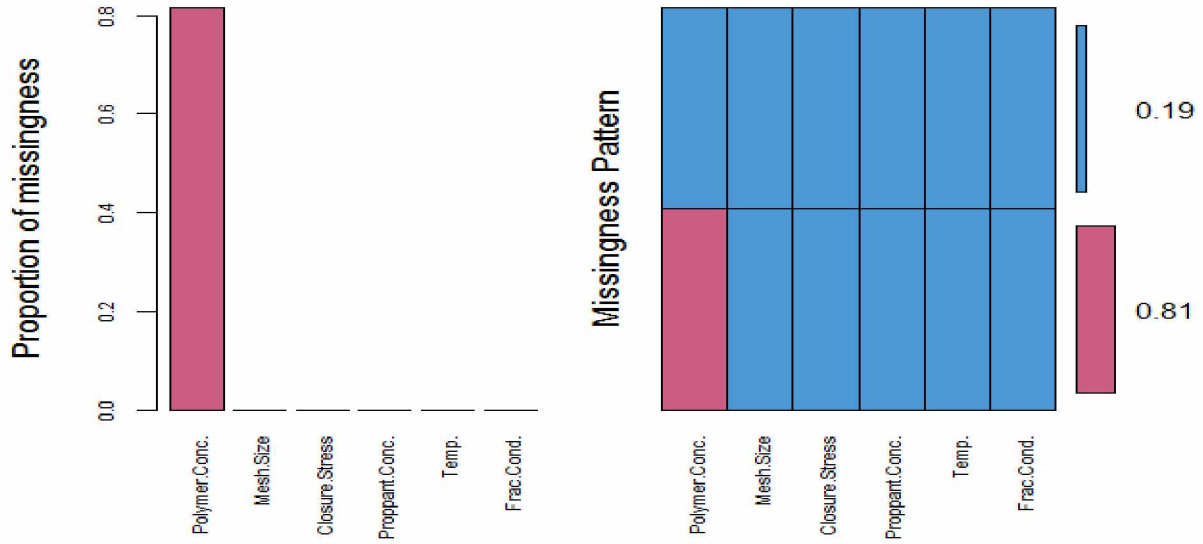

Figure 4-37: Proportion of missingness of polymer concentration in the dataset of Proppant Type-1

After two thousand sample runs, the first thousand sample runs are discarded for further analysis. The second thousand sample runs provide one thousand draws of regression model coefficients

Figure 4-38 shows the trace plot of all the model coefficients of second thousand sample runs. It shows the values that each parameter reached during the runtime. The plot clearly indicates that all the coefficients did not travel in any lengthy excursion away from the mean. There have been no discrepancies in the magnitudes of the coefficients and the chain has converged appropriately.

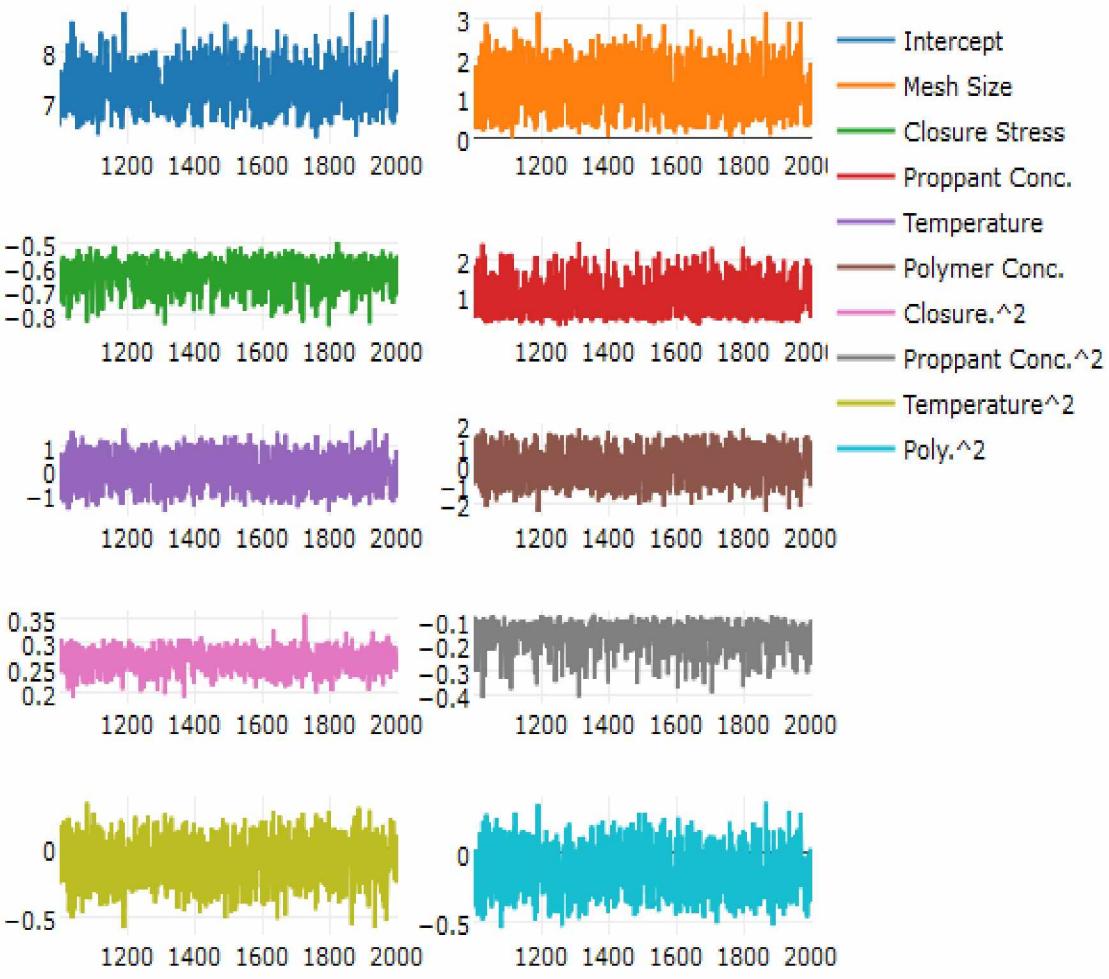

Figure 4-38: Trace plot of the model coefficients for second thousand sample runs for Proppant Type-1

Figure 4-39 shows the histograms of these model coefficients, which provide a clearer view to the actual magnitudes of all the coefficients. The plot also provides information about if any of the model coefficients have significantly resulted in zero or close to zero. If that happened then that particular model coefficient could be discarded. However, in **Figure 4-39,** the reddish color indicates the significant counts for the model coefficients. The figure suggests that no coefficients significantly occur at the magnitude of zero or close to zero.

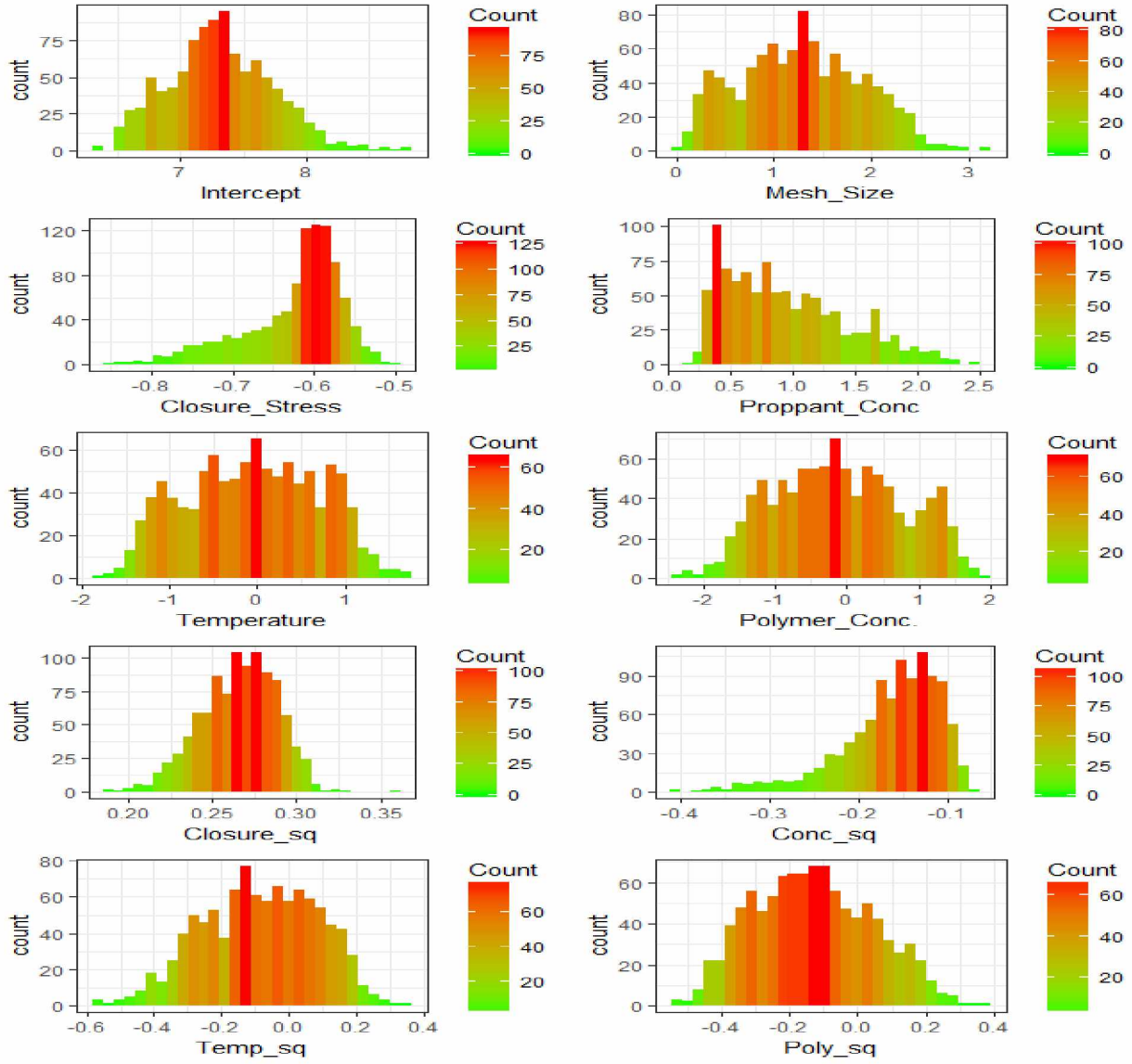

Figure 4-39: Histograms of model coefficients of all the predictors of Proppant Type-1 for second thousand sample runs

Finally, the median of all the model coefficients are selected as the ultimate conductivity model. The model is represented in **Equation 4.15**.

$$Fc = exp[7.30169 + (1.29819 * P_{ms}) - (0.61103 * \sigma_c) + (0.84021 * C_p) - (0.01733 * T) - (0.17320 * C_{pl}) + (0.26691 * {\sigma_c}^2) - (0.15161 * {C_p}^2) - (0.08523 * T^2) - (0.13015 * {C_{pl}}^2)] \quad (4.15)$$

4.3.2.2 Proppant Type-3

Figure 4-40 shows the proportion of missingness of polymer concentration. Around 24% of the data are available and the remaining 76% are to be imputed using multiple imputation.

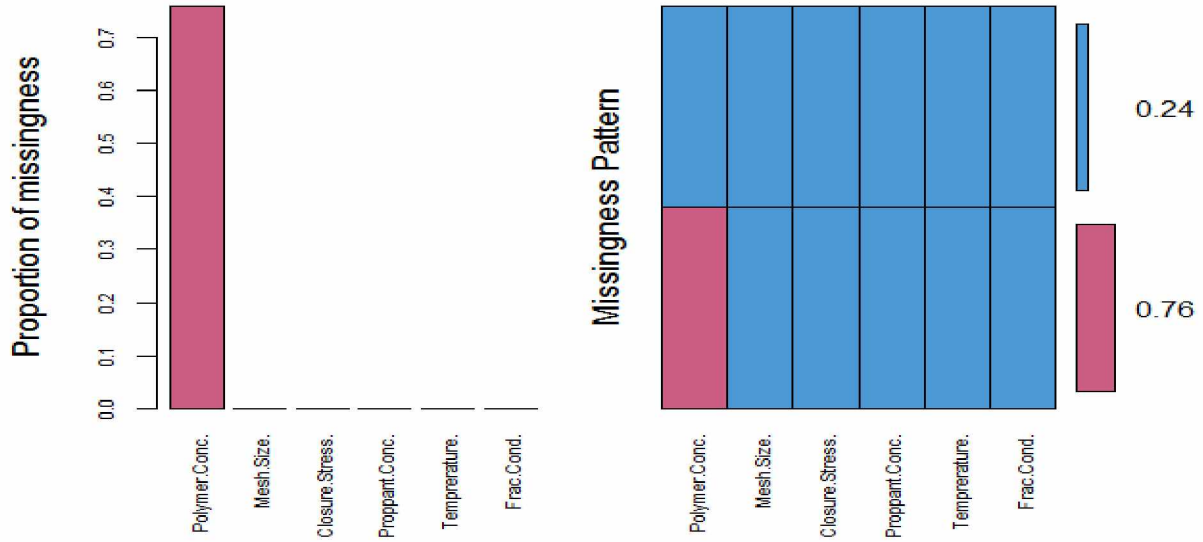

Figure 4-40: Proportion of missingness of polymer concentration in the dataset of Proppant Type-3

Figure 4-41 shows the trace plot of all the model coefficients of second thousand sample runs. This plot also suggests no inconsistencies in the values during the runtime. Precisely, the chain has converged well.

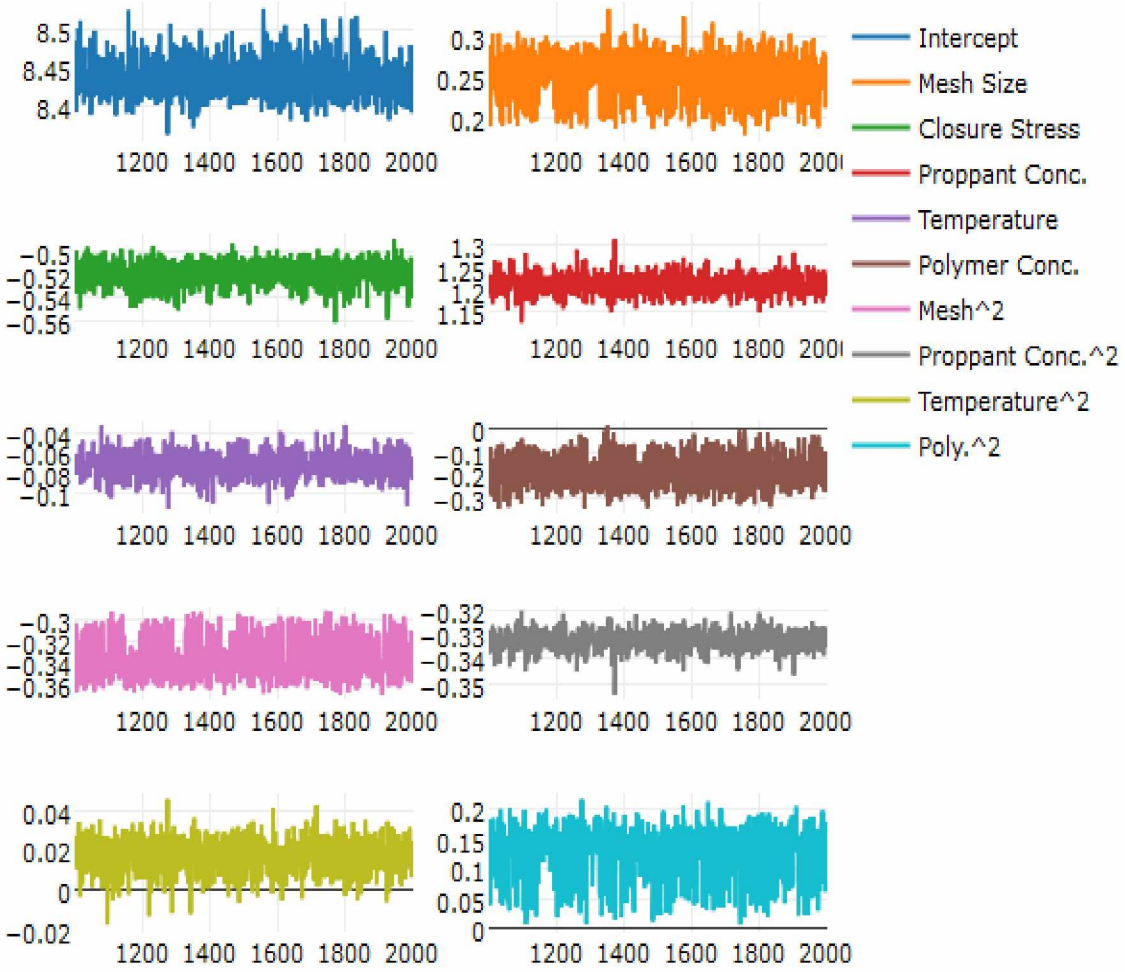

Figure 4-41: Trace plot of the model coefficients for second thousand sample runs for Proppant Type-3

Figure 4-42 shows the histograms of all the model coefficients for second thousand sample runs. No values of any of the coefficients were zero and therefore did not need to be discarded from the ultimate model.

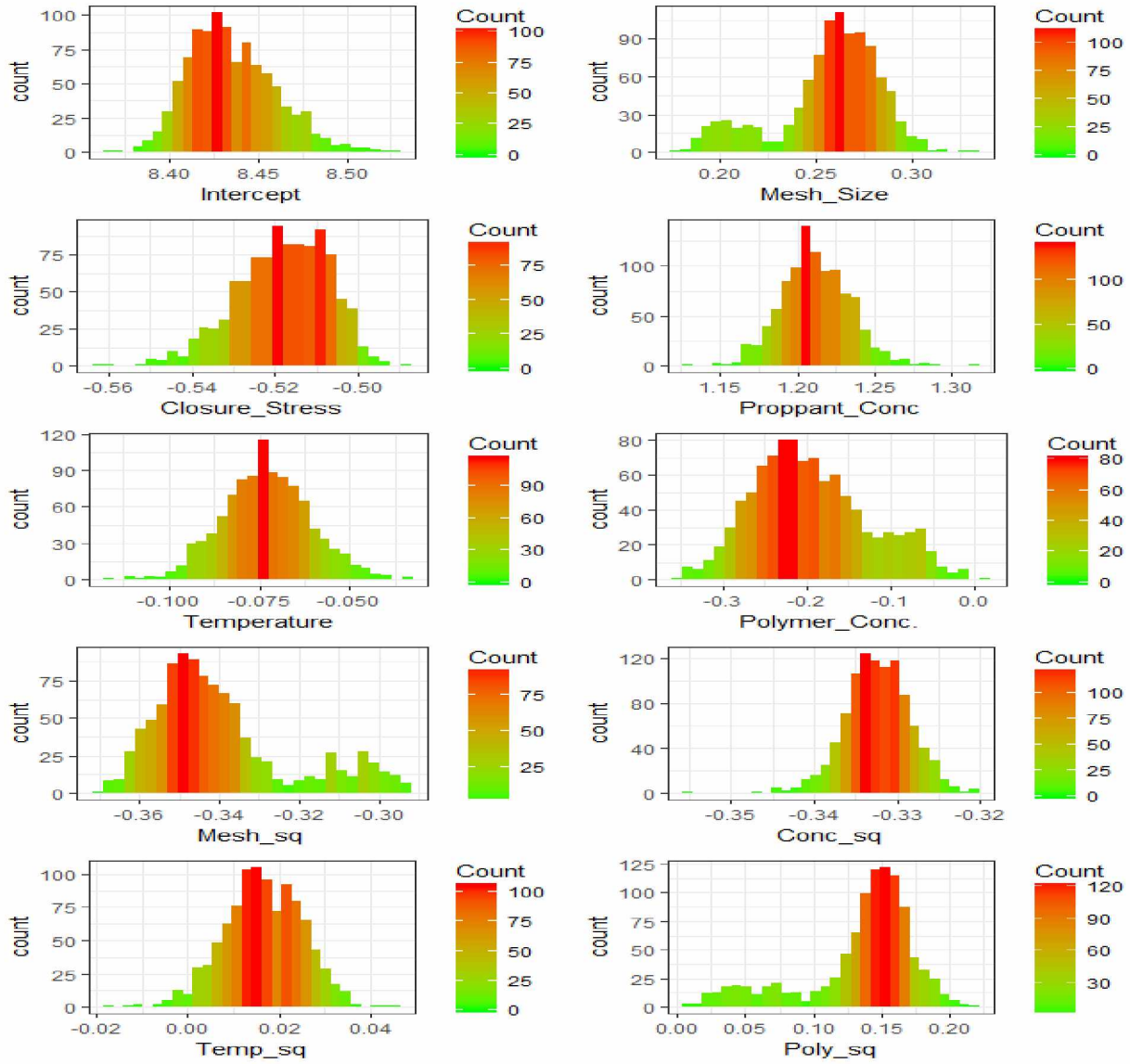

Figure 4-42: Histograms of model coefficients of all the predictors of Proppant Type-3 for second thousand sample runs

The final model using all medians of all the coefficients is given in the following **Equation 4.16**.

$$Fc = exp[8.43251 + (0.26211 * P_{ms}) - (0.51737 * \sigma_c) + (1.21181 * C_p) - (0.07340 * T) - (0.20376 * C_{pl}) - (0.34369 * {P_{ms}}^2) - (0.33233 * {C_p}^2) + (0.01651 * T^2) + (0.14597 * {C_{pl}}^2)] \quad (4.16)$$

4.3.3 Model Validation

4.3.3.1 Proppant Type-1

Both of the models are validated with 30% validation dataset of Proppant Type-1. **Figure 4-43** shows the validation of both the categorical imputation model (left plot) and the multiple imputation model (right plot) of Proppant Type-1. The prediction interval for the multiple imputation model (right plot) seems to be much wider than the categorical imputation model (left plot), but the actual conductivities are much closer to the prediction baseline for multiple imputation model. The mean squared prediction errors (MSPE) of are reported as 3.13 and 2.89 for the categorical imputation model and the multiple imputation model respectively. Therefore, the multiple imputation model for Proppant Type-1 seems to provide better predictions with new data than the categorical imputation model.

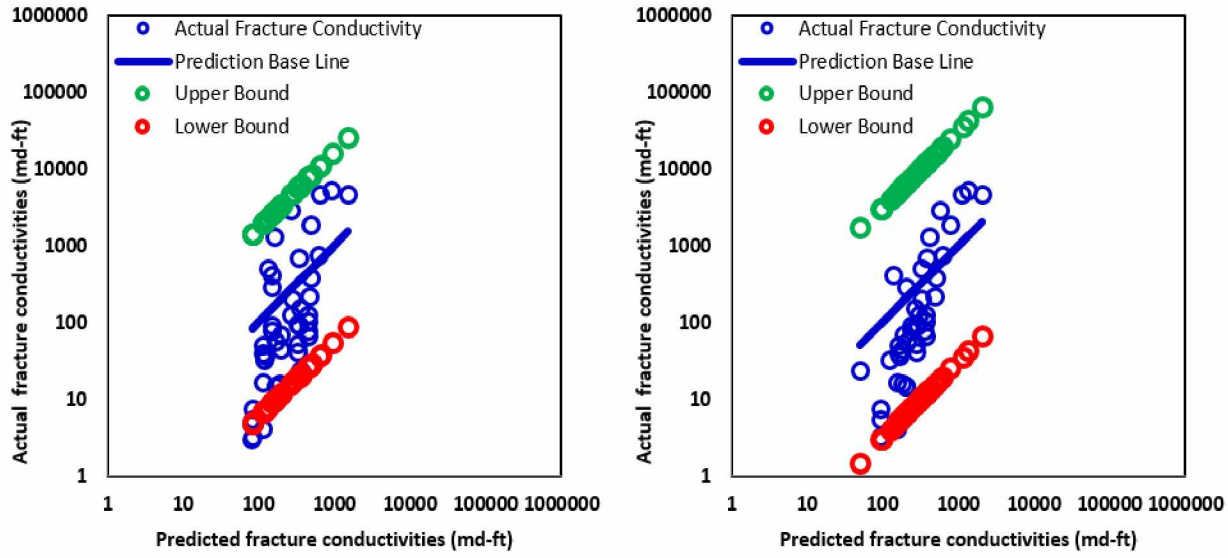

Figure 4-43: Graphical representation of predictions for both models of categorical imputation (left plot) and multiple imputation for Proppant Type-1 on 30% validation data with 95% prediction interval

4.3.3.2 Proppant Type-3

Figure 4-44 shows the model validation for both the imputation models of Proppant Type-3. The prediction interval for the categorical imputation model (left plot) is more tapered than the multiple imputation model (right plot), but similar to Proppant type-1, as more actual conductivities are inside the prediction interval envelope. The mean squared prediction errors (MSPE) for categorical imputation and multiple imputation are 0.64 and 0.61 respectively. Therefore, the multiple imputation seems to be more efficient than the categorical imputation model in terms of predictions with new data.

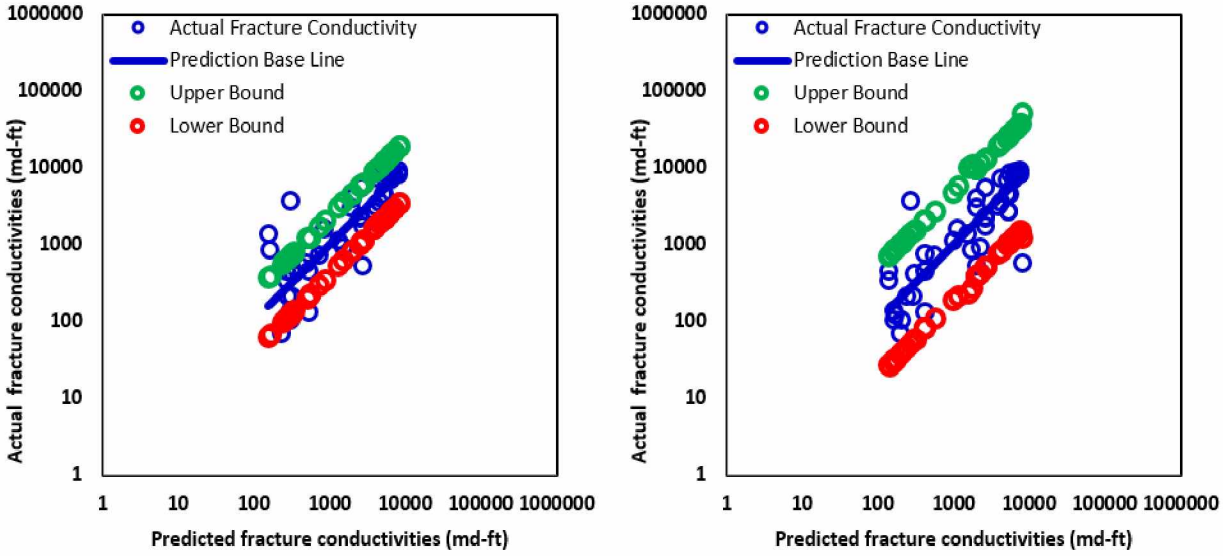

Figure 4-44: Graphical representation of predictions for both models of categorical imputation (left plot) and multiple imputation for Proppant Type-1 on 30% validation data with 95% prediction interval

4.4 Model Summary

Table 4-8 summarizes the correlations of all the five factor models of Proppant Type-1 and Proppant Type-3.

Table 4-8: Summary of the five models of Proppant Type-1 and Proppant Type-3

Proppant Type-1	
Fixed Effect Model	$Fc = exp[6.93165 + (1.26923 * P_{ms}) - (0.90890 * \sigma_c) + (3.44982 * C_p) - (0.10642 * T) - (0.08941 * C_{pl}) + (0.16575 * {\sigma_c}^2) - (1.90421 * {C_p}^2) + (0.05304 * T^2) - (0.40768 * {C_{pl}}^2)]$
Mixed Effect Model	$Fc = exp[6.91425 + (1.27017 * P_{ms}) - (0.90784 * \sigma_c) + (3.45732 * C_p) - (0.13163 * T) - (0.08775 * C_{pl}) + (0.16459 * {\sigma_c}^2) - (1.90147 * {C_p}^2) + (0.05597 * T^2) - (0.39348 * {C_{pl}}^2)]$
Categorical Imp. Model	$Fc = exp[7.36743 + (0.83192 * P_{ms}) - (0.82860 * \sigma_c) + (0.74121 * C_p) - (0.01412 * T) + (0.61678 * C_{pl}) + (0.26926 * {\sigma_c}^2) - (0.05886 * {C_p}^2) - (0.15247 * T^2) - (0.56102 * {C_{pl}}^2)]$
Multiple Imp. Model	$Fc = exp[7.30169 + (1.29819 * P_{ms}) - (0.61103 * \sigma_c) + (0.84021 * C_p) - (0.01733 * T) - (0.17320 * C_{pl}) + (0.26691 * {\sigma_c}^2) - (0.15161 * {C_p}^2) - (0.08523 * T^2) - (0.13015 * {C_{pl}}^2)]$
Proppant Type-3	
Fixed Effect Model	$Fc = exp[8.22287 + (0.78182 * P_{ms}) - (0.86664 * \sigma_c) + (1.70340 * C_p) - (0.05282 * T) - (0.03940 * C_{pl} - (0.87666 * {P_{ms}}^2) - (0.61180 * {C_p}^2) + (0.33036 * T^2) + (0.07949 * {C_{pl}}^2)]$
Mixed Effect Model	$Fc = exp[8.55960 + (1.20848 * P_{ms}) - (0.73940 * \sigma_c) + (1.51310 * C_p) - (0.66332 * T) - (0.22407 * C_{pl} - (1.10205 * {P_{ms}}^2) - (0.55294 * {C_p}^2) + (0.23563 * T^2) + (0.11727 * {C_{pl}}^2)]$
Categorical Imp. Model	$Fc = exp[8.54424 + (0.30032 * P_{ms}) - (0.53578 * \sigma_c) + (0.81348 * C_p) - (0.06157 * T) - (0.07549 * C_{pl}) - (0.30278 * {P_{ms}}^2) - (0.28091 * {C_p}^2) - (0.03320 * T^2) - (0.11691 * {C_{pl}}^2)]$
Multiple Imp. Model	$Fc = exp[8.43251 + (0.26211 * P_{ms}) - (0.51737 * \sigma_c) + (1.21181 * C_p) - (0.07340 * T) - (0.20376 * C_{pl}) - (0.34369 * {P_{ms}}^2) - (0.33233 * {C_p}^2) + (0.01651 * T^2) + (0.14597 * {C_{pl}}^2)]$

4.5 Comparison: Fixed vs. Mixed vs. Imputation Model

This section has three subsections. Each section includes a model comparison between all the models of five factors developed in this research for Proppant Type-1 and Proppant Type-3. In the first sub section, a summary of the model comparison is provided in terms of mean squared error (MSE) of model and of predictions. The next sub section is comprised of graphical representations illustrating the model comparison along with a previous conductivity model. The final subsection includes model comparisons with available case histories.

4.5.1 Comparison: Mean Squared Error (MSE)

Table 4-9 and **Table 4-10** represent the model comparison using mean squared error (MSE) as a performance assessing measure for Proppant Type-1 and Proppant Type-3 respectively. **Table 4-9** shows that both MSE and MSPE are the least for the mixed effect model. However, it should be noted that the mean squared error (MSE) of the model can be higher if large amount of data points are used to develop the model. If only imputation models are considered with each other, it seems unreasonable to conclude that one of them is the better one, because the categorical imputation model has a lower mean squared error (MSE) but the mean squared prediction error (MSPE) is lower for the multiple imputation model. Nevertheless, the mixed effect model could be chosen as the better one in terms of mean squared error (MSE). Imputation models could provide better predictions with newer data because they are developed using a larger datasets than the fixed and mixed effect model. But then they also suffer from the fact that part of their "larger dataset" was invented by the rest of the dataset.

Table 4-9: Model comparison by mean squared error (MSE) for Proppant Type-1

Performance Assessing Measure	Fixed Effect Model	Mixed Effect Model	Categorical Imputation Model	Multiple Imputation Model
MSE	0.33	0.32	2.01	2.16
MSPE	0.48	0.48	3.13	2.89

Table 4-10 indicates that both of the imputation models are better than the fixed and mixed effect models because they have lower mean squared error (MSE) and mean squared prediction error (MSPE). Again, it is unreasonable to choose the better one between the imputation models because one of them has a lower mean squared error (MSE) of model and the other has a mean squared prediction error (MSPE).

Table 4-10: Model comparison by mean squared error (MSE) for Proppant Type-3

Performance Assessing Measure	Fixed Effect Model	Mixed Effect Model	Categorical Imputation Model	Multiple Imputation Model
MSE	0.79	0.63	0.17	0.34
MSPE	0.80	0.81	0.64	0.61

4.5.2 Comparison: Previous Conductivity Model

Barree et al. (2016) have published a paper providing a general correlation of fracture conductivity which is facilitated by the Stim-Lab Proppant Conductivity Consortium. The model developed in this study used datasets containing a wide range of proppant types and sizes. The main aim was to provide correlations of baseline conductivity to assess newer proppant types or sizes. However, the model was developed as a function of closure stress and proppant properties and concentration. Thus, the 30% validation data used to validate all the models of this research is used to predict fracture conductivity by using the correlation provided by Barree et al. (2016). The model has specified different regression coefficients for different proppants of each proppant type. To use the correlation with the validation data, these regression coefficients are averaged for one proppant type. For example, if different coefficients are provided for brown sand, white sand, then different sands are grouped as one and one averaged value is used for that particular proppant type. Mesh size is transformed from US sieve to micron according to the API specifications and later averaged for one proppant type. Comparisons for both Proppant Type-1 and Proppant type-3 with the developed models of this research and Barree et al. (2016) model are consecutively shown in **Figure 4-45** and **Figure 4-46**. The aim is to observe the predictive performance of each model on the same validation data.

Figure 4-45 shows that fixed and mixed effect models are the best predictive models among all of them, and they display almost similar predictions. All the actual conductivities are very close and well around the prediction base line for these two cases. On the other hand, both the imputation models appear to provide more scattered predictions. Additionally, the figure also indicates that the Barree et al. (2016) model has a tendency to provide underestimated predictions. Precisely, the fixed and mixed effect models appear to be the better ones for Proppant Type-1.

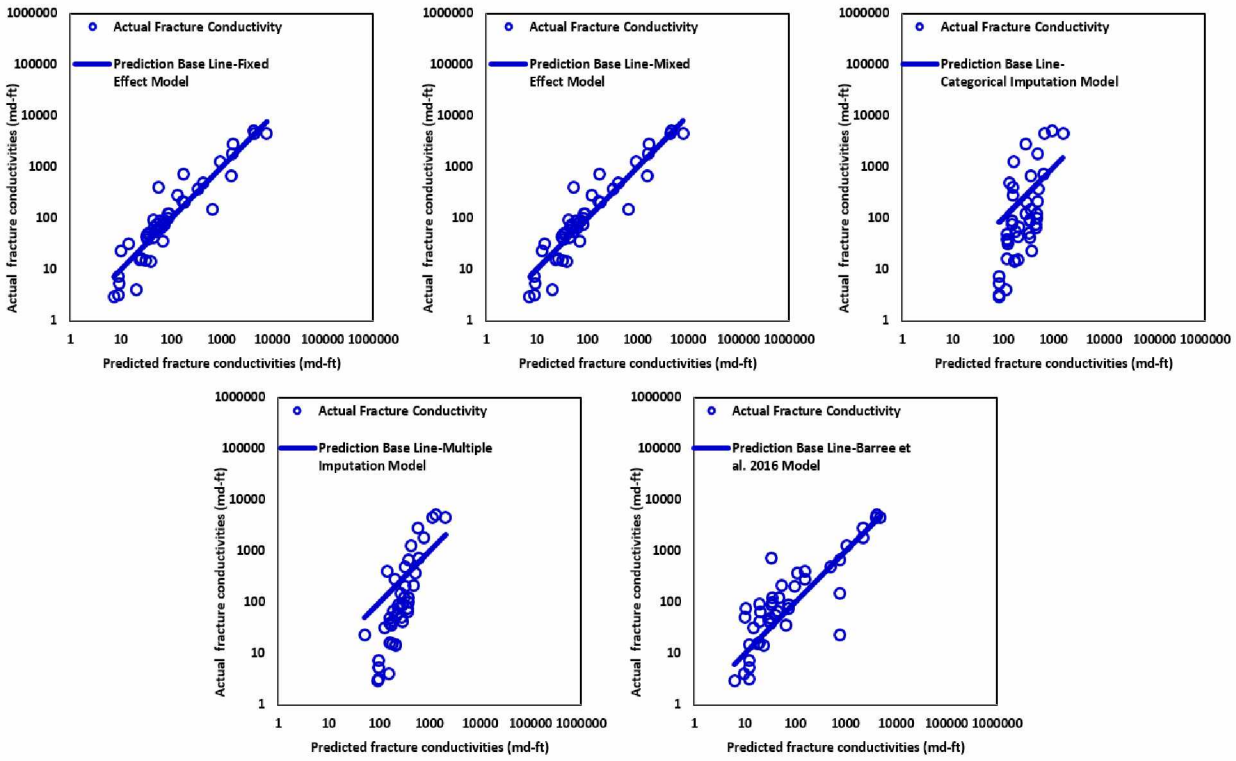

Figure 4-45: Graphical measures of model predictive performance for models of Proppant Type-1 used in this research, as well that of Barre et al. model Fracture Conductivity Model (Top left-fixed effect model, top middle-mixed effect model, top right-categorical imputation model, bottom left-multiple imputation model, bottom right-Barree et al. (2016) model)

Furthermore, **Figure 4-46** shows the comparison between the models of this research and Barree et al. (2016) model for Proppant Type-3. The figure clearly suggests that the multiple imputation model is the best one as the actual conductivities are very close to the prediction base line compared to the others. The categorical imputation model also seems to provide very good predictions. The predictions from fixed and mixed effect models look acceptable and identical to

each other. On the other hand, the Barree et al. (2016) model shows that the actual conductivities are more scattered from the prediction base line. Consequently, the models of Proppant Type-3 can be arranged in order of their predictive performance: multiple imputation model, categorical imputation model, fixed and mixed effect model, Barree et al. (2016) model.

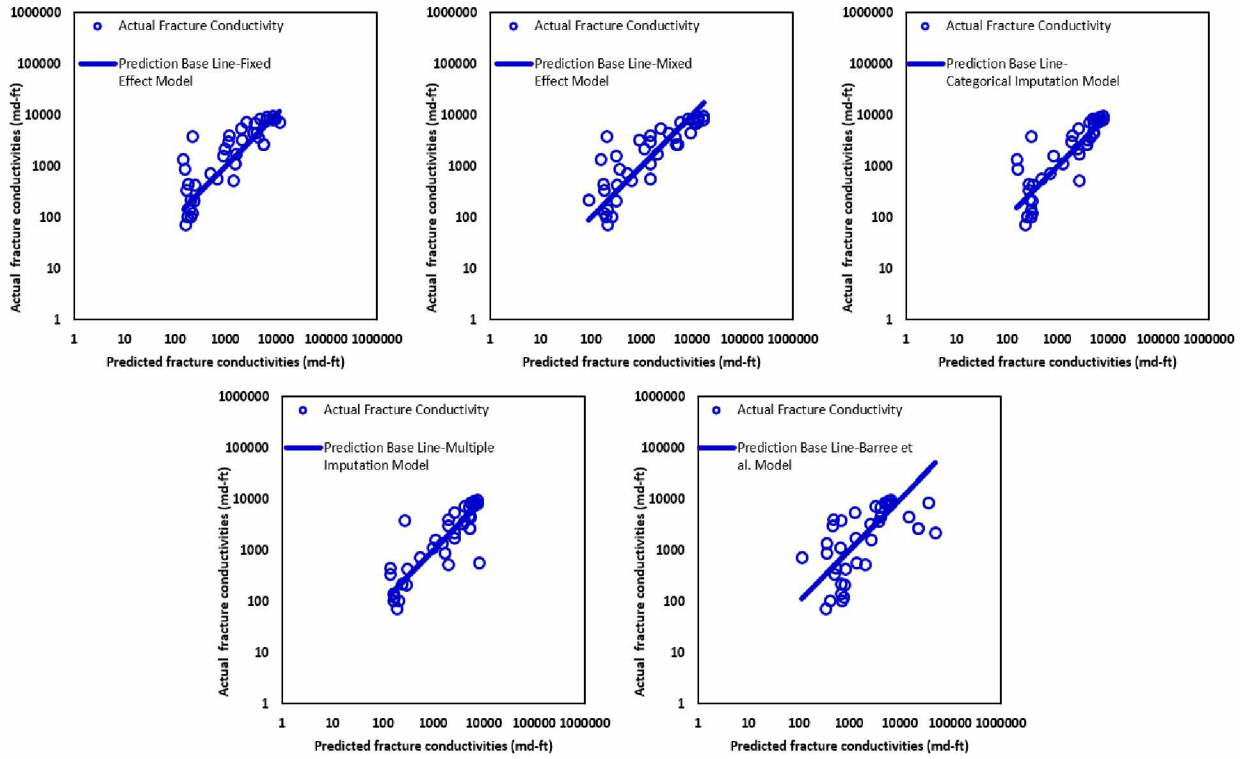

Figure 4-46: Graphical measures of model predictive performance for models of Proppant Type-3 used in this research, as well that of Barre et al. model Fracture Conductivity Model (Top left-fixed effect model, top middle-mixed effect model, top right-categorical imputation model, bottom left-multiple imputation model, bottom right-Barree et al. (2016) model)

In conclusion, it is not possible to suggest only one of the statistical approaches as more accurate for both types of proppant, although in general, fixed and mixed effect models appear to provide more decent predictions for both proppant types. However, these models are recommended for use in between a range of small magnitudes of independent variables while the ranges for using imputation models are wider. Clearer assumptions can be made from the results of case histories in the next section.

4.5.3 Comparison: Case Histories

The main aim of this section is to collect realistic data of the independent variables of this research and to use it in the models developed here. Subsequently, the predictions are then compared with the real conductivity of that particular case. All the case histories discussed in this section did not report any values of polymer concentration. However, an approximate value can be chosen based on the previous experimental investigations on proppant pack conductivity. Awoeleke et al. (2012) used polymer concentration of 10 lb/1000 gal and 30 lb/1000 gal for their experiments. Marpaung et al. (2008) varied the polymer concentration in between the range of 30-50 lb/1000 gal for their dynamic conductivity tests. Penny (1987) evaluated the dynamic leakoff characteristics of titanate crosslinked HP guar system using 40, 50, and 60 lb/1000 gal of polymer loadings. Therefore, polymer concentration of 40 lb/ 1000 gal is assumed for all the cases. More recent data that explicitly provides values of all the independent variables of this research is scarce. Therefore, the most recent data are from 2007 (Handren and Palisch, 2007) for this case study.

4.5.2.1 Case History I

Lindley (1983) investigated proppant placement in a hydraulic fracturing treatment "tail-in" technique at McAllen Ranch Field, Texas. The reservoir of McAllen Ranch Field has high temperatures but the amount is not reported in that research, so a temperature of 300^0F is assumed for this case. Data from two wells are used here for comparison: McAllen 16 and McAllen 17. Hydraulic fracture conductivity is referred to as the fracture flow capacity data in this research. This data is calculated using pressure build up test. For well 16, pressure build up tests were done on 02/10/1978 and on 04/26/1983. Pressure build up tests were also carried out twice for well 17 on 02/19/1980 and on 05/17/1982. For the hydraulic fracturing, treatment 20/40 bauxite is used as proppant. The models developed for ceramic proppants are used to predict the conductivity of these two wells. **Figure 4-47** and **Figure 4-48** show the predictions of the models for the case of pressure build up tests results of well 16 dated on 02/10/1978 and 04/26/1983 respectively. Similarly, **Figure 4-49** and **Figure 4-50** show results for the case of well 17 dated on 02/19/1980 and 05/17/1982 respectively. Red dots indicate the predictions and the blue error bands indicate the prediction intervals. In both **Figure 4-47** and **Figure 4-48**, the predictions from multiple imputation are closer to the red dotted horizontal line indicating the actual conductivity data. Undoubtedly, for well 16, the multiple imputation model provides the best predictions. For well 17, both the mixed effect model and the multiple imputation model seem to provide decent

predictions closer to actual conductivity line. Thus, the predictive performance of the multiple imputation model can be claimed as the best. Actual conductivities of both wells do fall inside the prediction interval for almost all the models without the predictions from categorical imputation for well 17.

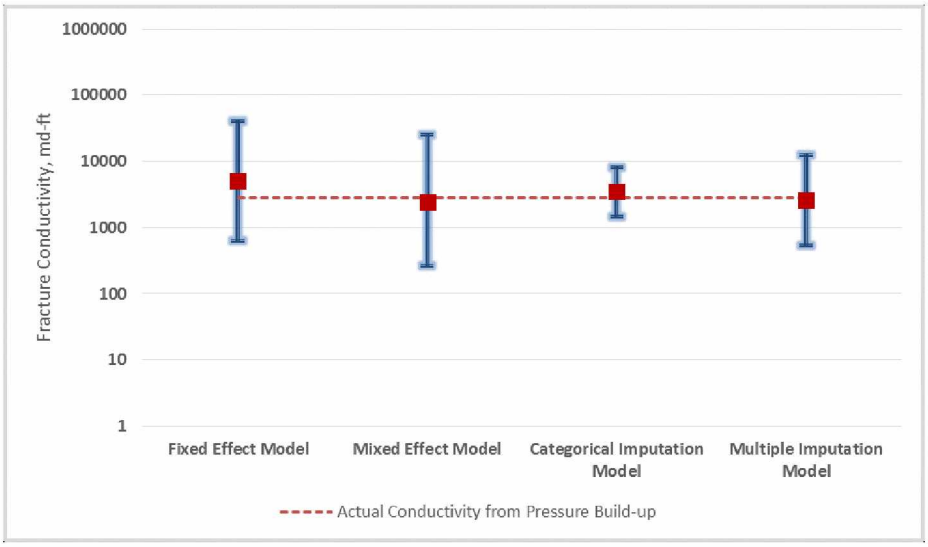

Figure 4-47: Comparison of the predictive performances of all the models developed for Proppant Type-3 in this research with well 16 (dated on 02/10/1978) of McAllen Field, Texas

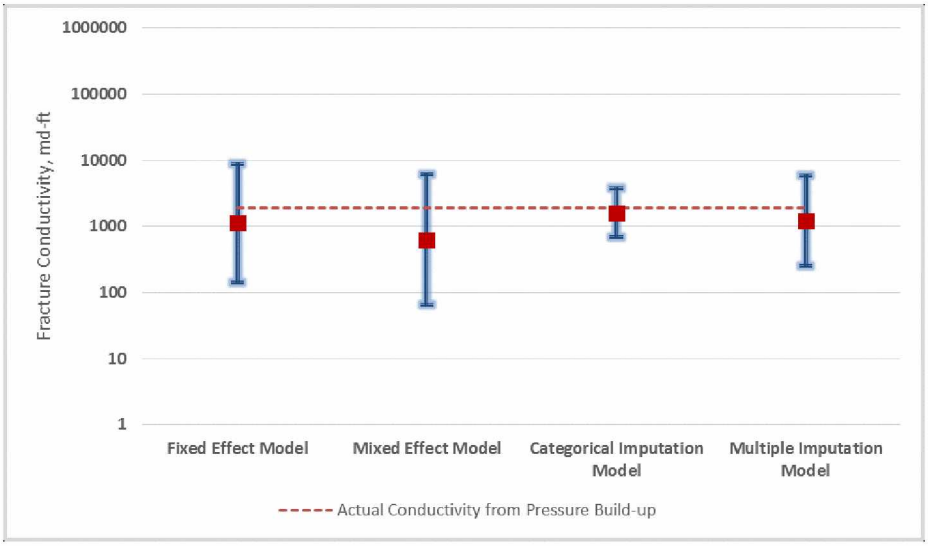

Figure 4-48: Comparison of the predictive performances of all the models developed for Proppant Type-3 in this research with well 16 (dated on 04/26/1983) of McAllen Field, Texas

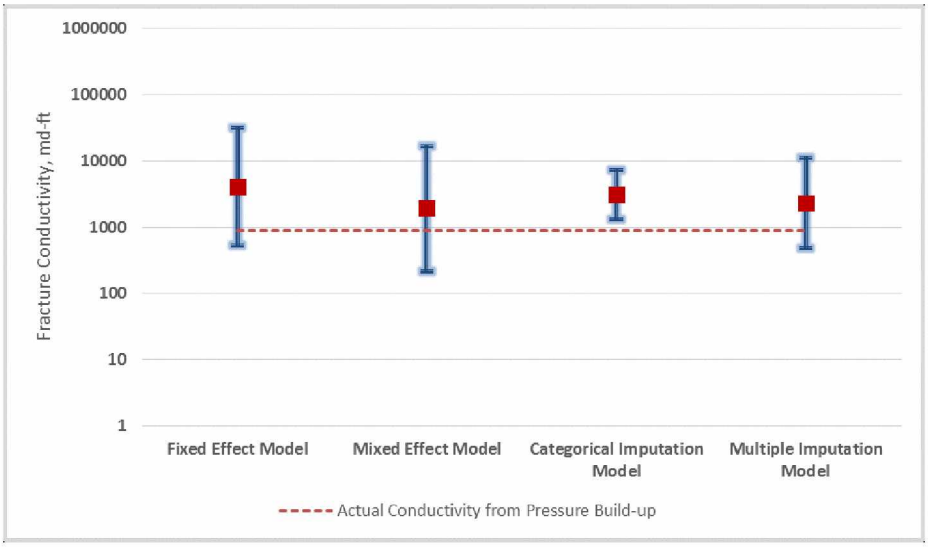

Figure 4-49: Comparison of the predictive performances of all the models developed for Proppant Type-3 in this research with well 17 (dated on 02/19/1980) of McAllen Field, Texas

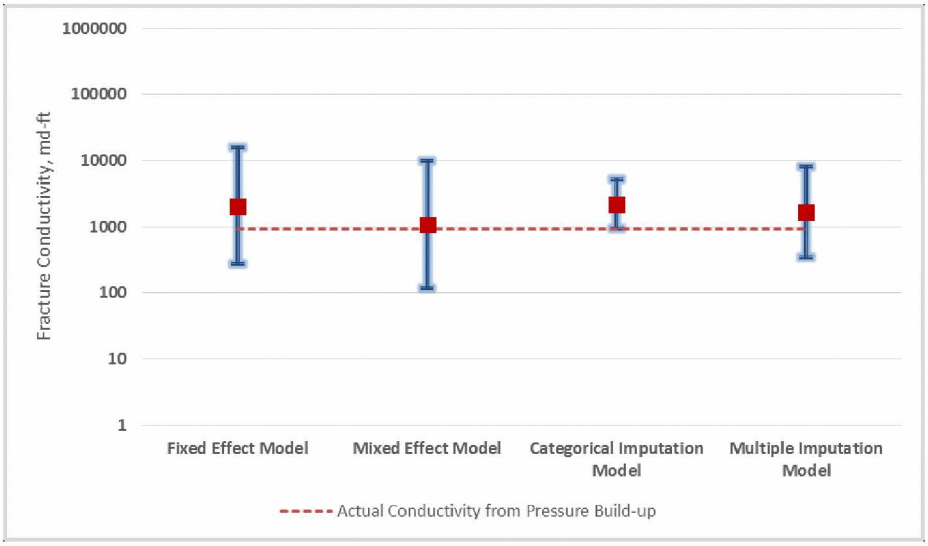

Figure 4-50: Comparison of the predictive performances of all the models developed for Proppant Type-3 in this research with well 17 (dated on 05/17/1982) of McAllen Field, Texas

4.5.2.1 Case History II

Palisch et al. (2007) reported realistic hydraulic fracture conductivity using PredictK (Stim Lab Proppant and Fluid Consortia, 1997-2006) for Wamsutter Gas Field, Wyoming. The PredictK software allowed them to identify the realistic effects of the individual factors on proppant pack conductivity. The gas field includes around 1300 wells, and over 150 wells were fractured from 2002-2004. PredictK assumed a proppant concentration of 0.5 lb/ft^2. Bottomhole temperature and closures stress are reported as 225^0F and 6000 psi respectively. PredictK reports baseline conductivity for white sand, RCS, ELWC, and ISP. No mesh size is reported in the paper. So, a range of mesh sizes is used for each proppant type and then the predicted conductivity is averaged. The average predicted conductivities from all models of Proppant Type-1 are compared with the reported conductivity of white sand. Similarly, the averaged predicted conductivities of Proppant Type-3 are compared with the averaged conductivity of ELWC and ISP. **Figure 4-51** and **Figure 4-52** show the results of the model comparison. Undeniably, the multiple imputation models for both proppant types are the best in terms of predictive performance. The mixed effects models also provide decent predictions for both types of proppant and can be referred as the second best. All actual conductivities are inside the prediction intervals.

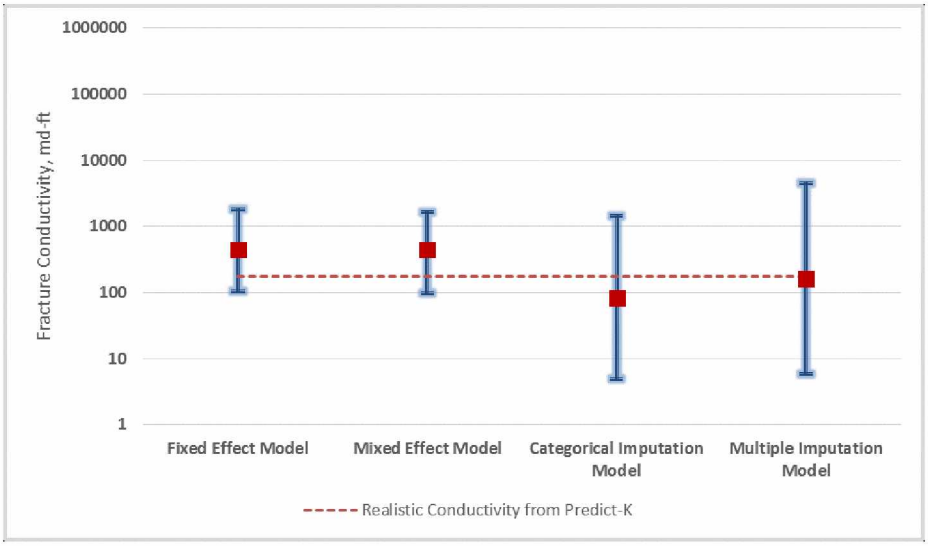

Figure 4-51: Comparison of the predictive performances of all the models developed for Proppant Type-1 in this research with realistic conductivity of Wamsutter Field, Wyoming from PredictK

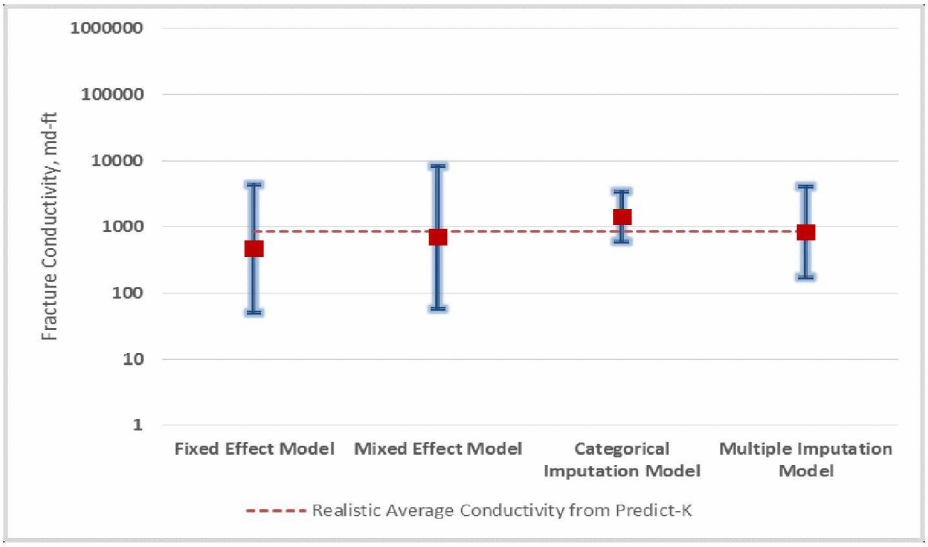

Figure 4-52: Comparison of the predictive performances of all the models developed for Proppant Type-3 in this research with realistic conductivity of Wamsutter Field, Wyoming from PredictK

4.5.2.3 Case History III

Handren and Palisch (2007) illustrated a case history of hybrid slickwater-fracture designs of Cotton Valley Taylor completions of east Texas. The authors simulated the actual well conditions

and determined proppant performance at temperature and stress conditions. The actual conditions were maintained for 45 days to analyze the loss in fracture conductivities. The authors conducted tests using different proppants. Results from 40/70 sand, 20/40 ELWC, and 30/50 ELWC proppants are used here to carry out the model comparison. Temperature, closure stress, and proppant concentration are reported as 250^0F, 6000 psi, and 2 lb/ft^2. The predictions for 40/70 sand, 20/40 ELWC, and 30/50 ELWC are displayed in **Figure 4-53**, **Figure 4-54,** and **Figure 4-55** respectively. The greenish dotted horizontal line represents the conductivity after 1.5 months. **Figure 4-53** suggests that predictions from fixed, mixed, and multiple imputations are close to each other, although the multiple imputation model predicts a little better than the other two. All the predictions are not close enough to the actual conductivity presented by red dotted line. The reduced conductivity after 1.5 months, however, is much closer to the predictions. In **Figure 4-54**, the predicted conductivity from multiple imputation appears to be the closest to the actual conductivity.

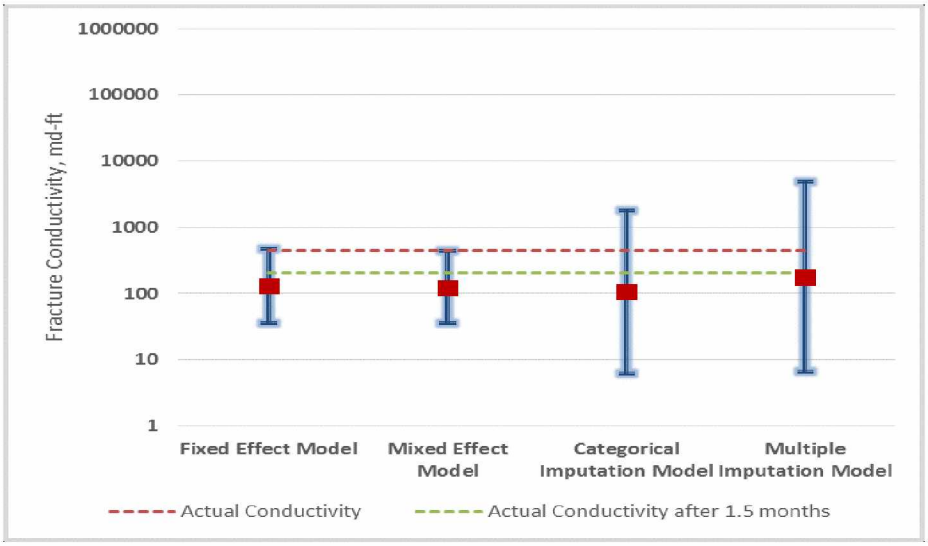

Figure 4-53: Comparison of the predictive performances of all the models developed for Proppant Type-1 in this research with 40/70 sand used in Cotton Valley Taylor, Texas

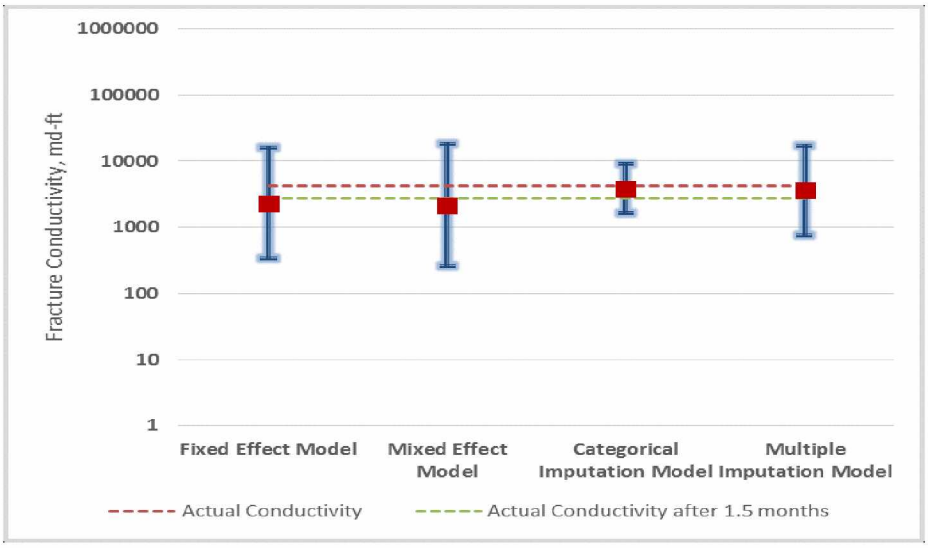

Figure 4-54: Comparison of the predictive performances of all the models developed for Proppant Type-3 in this research with 20/40 ELWC used in Cotton Valley Taylor, Texas

Last but not least, **Figure 4-55** shows that the predictions using the fixed and the mixed models for 30/50 ELWC seem to be well deviated from the actual conductivity. The categorical imputation and the multiple imputation models provide better predictions.

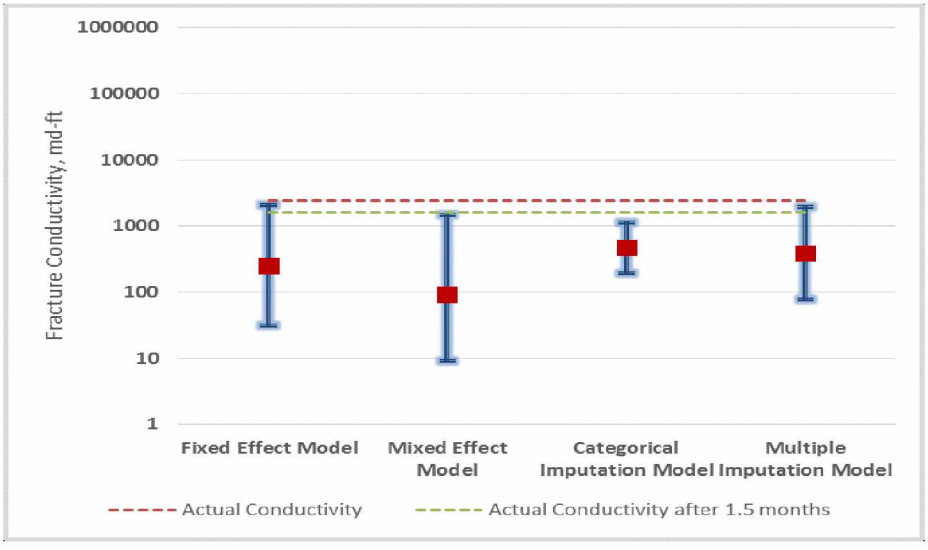

Figure 4-55: Comparison of the predictive performances of all the models developed for Proppant Type-3 in this research with 30/50 ELWC used in Cotton Valley Taylor, Texas

Results from all the case histories suggest that the multiple imputation models for both proppant types showed better predictions at real conditions because the ranges of the data that were used to develop the model are broader than the other models. The mixed model is also recommended to use for predictions. However, the engineers who have access to wide range of recent fracturing data can perform further validation of the models.

Chapter 5 CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

This study presented several empirical models of hydraulic fracture conductivity using different statistical tools. The study was mainly focused on five factors (proppant particle size, closure stress, proppant concentration, temperature, and polymer concentration) affecting conductivity. Advanced statistical tools were employed on datasets of two proppant types: sand and ceramic proppants. However, the other models developed here might be useful but yet to be investigated extensively. The following conclusions are made based on the entire study:

1. Reservoir temperature and polymer concentration are statistically least significant amongst all. However, it is important to consider these two factors for future predictions. Proppant particle size, closure stress, and proppant concentration are the most significant factors.
2. Increase in proppant particle size and proppant concentration has positive effects on hydraulic fracture conductivity. Increase in closure stress, temperature, and polymer concentration affects the hydraulic fracture conductivity adversely. However, the changes are observed within the limits of the studied data.
3. The mixed effect models for both proppant types (Proppant Type-1 and Proppant Type-3) provide the least mean squared error (MSE).
4. Based on the comparison with the previous hydraulic fracture conductivity model (Barree et al., 2016) and realistic data, multiple imputation models for both proppant types are more dependable for predictions than the other models.
5. Therefore, the mixed and the multiple imputations models presented here can be used by the engineers for future predictions.
6. Multiple imputation models are recommended because the ranges of the input data into the model are wide.
7. Engineers can use the correlations as input in a simulator that can be used to predict fracture conductivity. However, it is important to keep in mind that the models are recommended to use while the new predictors data are inside the ranges provided in **Appendix B**.

The limitations of this study are stated in the following:

1. Multiple imputation and mixed models are only valid for sand and ceramic proppants. No effective correlation is provided which is valid with resin coated sand.
2. This study has provided several models to predict fracture conductivity but many of them were neither compared with any previous hydraulic conductivity model nor with realistic data. The fixed model of six factors was statistically adequate and could provide good predictions.

5.2 Recommendations

The fixed model of six factors can be further investigated by employing mixed effect modeling and statistical imputation. The study only includes two statistical imputation approaches. Therefore, other imputation methods can also be employed in future to determine the best approach. The datasets used in this study include both short-term and long-term conductivity data. Thus, adding time as an independent variable may provide more precise correlations of hydraulic fracture conductivity.

REFERENCES

Allison, P. 2015. Imputation by Predictive Mean Matching: Promise & Peril. *Statistical Horizon,* March 5, 2015, https://statisticalhorizons.com/predictive-mean-matching (accessed 21 September, 2016).

Anderson, R.W., Cooke, C.E., and Wendorff, C.L. 1989. Propping Agents and Fracture Conductivity. *Recent Advances in Hydraulic Fracturing*, ed. Gidley, J. L., Holditch, S.A., Nierode, D. E. et al. Chap. 6, 109-130. Richardson, TX: SPE Monograph Series Vol **2**, Society of Petroleum Engineers.

API RP61. 1989. *Recommended Practices for Evaluating Short Term Conductivity*. Washington, DC: API.

Aven, K. N., Weaver, J. D., and Tang, T. 2013. Long-Term Dynamic Flow Testing of Proppants. Society of Petroleum Engineers. Presented at SPE International Symposium on Oilfield Chemistry, The Woodlands, Texas, 8-10 April. SPE-164082-MS.

https://doi.org/10.2118/164082-MS.

Awoleke, O. O., Romero, J. D., Zhu, D. et al. 2012. Experimental Investigation of Propped Fracture Conductivity in Tight Gas Reservoirs Using Factorial Design. Presented at SPE Hydraulic Fracturing Technology Conference, The Woodlands, Texas, 6-8 February. SPE-151963-MS.

https://doi.org/10.2118/151963-MS.

Awoleke, O. O., Zhu, D., and Hill, A. D. 2016. New Propped-Fracture-Conductivity Models for Tight Gas Sands. *SPE Journal* **21** (05): 1508-1517. SPE-179743-PA.

https://doi.org/10.2118/179743-PA.

Barree, R. D., Miskimins, J. L., Conway, M. W. et al. 2016. Generic Correlations for Proppant Pack Conductivity. Presented at SPE Hydraulic Fracturing Technology Conference, The Woodlands, Texas, 9-11 February. SPE-179135-MS.

https://doi.org/10.2118/179135-MS.

Bates, D., Maechler, M., Bolker, B. et al. 2015. Fitting Linear Mixed-Effects Models Using lme4 . *Journal of Statistical Software* **67** (1): 1-48.

https://doi.org/10.18637/jss.v067.i01.

Buckingham, E. 1914. On Physically Similar Systems; Illustrations of the Use of Dimensional Equations. Phys. Rev. IV(4): 345-376.

http://dx.doi.org/10.1103/PhysRev.4.345.

Buuren, S. V. and Oudshoorn, K. G. 2011. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* **45** (3): 1-67.

http://www.jstatsoft.org/v45/i03/.

Cook, R. D. 1977. Detection of Influential Observations in Linear Regression. Technometrics. *American Statistical Association* **19** (): 15-18.

https://doi.org/10.2307%2F1268249.

Cooke, C. E. 1973. Conductivity of Fracture Proppants in Multiple Layers. *SPE Journal of Petroleum Technology* **25** (09): 1101-1107. SPE-4117-PA.

https://doi.org/10.2118/4117-PA.

Cooke, C. E. 1975. Effect of Fracturing Fluids on Fracture Conductivity. *SPE Journal of Petroleum Technology* **27** (10): 1273-1282. SPE-5114-PA.

https://doi.org/10.2118/5114-PA.

Coulter, G. R., and Wells, R. D. 1972. The Advantages of High Proppant Concentration in Fracture Stimulation. *SPE Journal of Petroleum Technology* **24** (06): 643-650. SPE-3298-PA. https://doi.org/10.2118/3298-PA.

Cutler, R. A., Enniss, D. O., Jones, A. H. et al. 1985. Fracture Conductivity Comparison of Ceramic Proppants. *SPE Journal* **25** (02): 157-170. SPE-11634-PA.

https://doi.org/10.2118/11634-PA.

Duenckel, R., Moore, N., O'Connell, L. et al. 2016. The Science of Proppant Conductivity Testing-Lessons Learned and Best Practices. Presented at SPE Hydraulic Fracturing Technology Conference, The Woodlands, Texas, 9-11 February. SPE-179125-MS.

https://doi.org/10.2118/179125-MS.

Fox, J. and Weisberg, S. 2011. An {R} Companion to Applied Regression, Second Edition. Thousand Oaks CA: Sage.

http://socserv.socsci.mcmaster.ca/jfox/Books/Companion.

Fredd, C. N., McConnell, S. B., Boney, C. L.et al. 2001. Experimental Study of Fracture Conductivity for Water-Fracturing and Conventional Fracturing Applications. *SPE Journal* **6** (03): 288-298. SPE-74138-PA.

https://doi.org/10.2118/74138-PA.

Geman, S., and Geman, D. 1984. Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6** (06):721-741.

https://doi.org/10.1109/TPAMI.1984.4767596.

Gidley, J. L., Hoditch, S. A., Nierode, D. E. et al. 1989. *Recent Advances in Hydraulic Fracturing*. TX: Society of Petroleum Engineers.

Gilks W. R. 2005. Markov Chain Monte Carlo. *Wiley Online Library,* 15 July, 2005, http://onlinelibrary.wiley.com/doi/10.1002/0470011815.b2a14021/abstract (accessed 01 Sepetember, 2016).

Grinstead, C. M., and Snell, J. L. 1997. *Markov Chains. Introduction to Probability*, 2nd revised edition, Chap. 11, 405-452. Providence, RI: American Mathematical Society.

GY Ceramic Proppant Co. Ltd. 2012. http://www.ceramic-proppants.com/about-ceramic-proppant.html (accessed 30 December, 2016).

Handren, P. J., and Palisch, T. T. 2007. Successful Hybrid Slickwater Fracture Design Evolution--An East Texas Cotton Valley Taylor Case History. Presented at SPE Annual Technical Conference and Exhibition, Anaheim, California, 11-14 November. SPE-110451-MS.

https://doi.org/10.2118/110451-MS.

Hill, A. D., Pournik, M., Zou, C. et al. 2007. Small-Scale Fracture Conductivity Created by Modern Acid Fracture Fluids. SPE Hydraulic Fracturing Technology Conference, College Station, Texas, 29-31 January. SPE-106272-MS.

https://doi.org/10.2118/106272-MS.

John, A. O., Joel, O.F., and Chukwuma, F.O. 2017. Evaluating the Effect of Temperature and Breaker Concentration on Degradation of Cross linked Hydroxypropyl Guar Fluid. *American Journal of Engineering Research* **6** (6): 34-40. e-ISSN: 2320-0847 p-ISSN : 2320-0936.

Johnson, JW. 2000. A Heuristic Method for Estimating the Relative Weight of Predictor Variables in Multiple Linear Regression. *Multivariate Behavioral Research* **35** (01): 1-19. PMID: 26777229.

http://dx.doi.org/10.1207/S15327906MBR3501_1.

Kabacoff, R. I. (2015). *R in Action, Data Analysis and Graphics with R,* 2nd Edition. Shelter Island, NY: Manning Publications Co.

Kainer, C., Guerra, D., Zhu, D. et al. 2017. A Comparative Analysis of Rock Properties and Fracture Conductivity in Shale Plays. Presented at SPE Hydraulic Fracturing Technology Conference and Exhibition, The Woodlands, Texas, 24–26 January. SPE-184877-MS.

https://doi.org/10.2118/184877-MS.

Kakkar, P. 2016. Experimental Study of the Effect of Stress and Fluid Sensitivity on Propped and Un-propped Fracture Conductivity in Preserved Reservoir Shale, The University of Texas at Austin, Austin, TX. (May, 2016).

http://hdl.handle.net/2152/39117.

Kamenov, A. N. 2013. The Effect of Proppant Size and Concentration on Hydraulic Fracture Conductivity in Shale Reservoirs, Texas A & M University, College Station, TX. (May, 2013).

http://hdl.handle.net/1969.1/149386.

Li, K., Gao, Y., Lyu, Y. et al. 2015. New Mathematical Models for Calculating Proppant Embedment and Fracture Conductivity. *SPE Journal* **20** (03): 496-507. SPE-155954-PA.

https://doi.org/10.2118/155954-PA.

Lindley, B. W. 1983. An Investigation of a High-Strength Proppant Tail-In at McAllen Ranch Field. Presented at SPE Annual Technical Conference and Exhibition, San Francisco, California, 5-8 October. SPE-11935-MS.

https://doi.org/10.2118/11935-MS.

Little, R. J. A. and Rubin D. B. 2002. *Statistical Analysis with Missing Data,* 2nd edition. Hoboken, New Jersey: John Wiley & Sons.

Lumley, T. 2017. Based on Fortran code by Miller, A. leaps: Regression Subset Selection. R pac kage version 3.0.

https://CRAN.R-project.org/package=leaps.

Marpaung, F., Chen, F., Pongthunya, P. et al. 2008. Measurement of Gel Cleanup in a Propped Fracture With Dynamic Fracture Conductivity Experiments. Presented at SPE Annual Technical Conference and Exhibition, Denver, Colorado, 21-24 September. SPE-115653-MS.

https://doi.org/10.2118/115653-MS.

McDaniel, B. W. 1986. Conductivity Testing of Proppants at High Temperature and Stress. Presented at SPE California Regional Meeting, Oakland, California, 2-4 April. SPE-15067-MS.

https://doi.org/10.2118/15067-MS.

McDaniel, B. W. 1987. Realistic Fracture Conductivities of Proppants as a Function of Reservoir Temperature. Presented at Low Permeability Reservoirs Symposium, Denver, Colorado, 18-19 May. SPE-16453-MS.

https://doi.org/10.2118/16453-MS.

McGinley, M., Zhu, D., and Hill, A. D. 2015. The Effects of Fracture Orientation and Elastic Property Anisotropy on Hydraulic Fracture Conductivity in the Marcellus Shale. Presented at SPE Annual Technical Conference and Exhibition, Houston, Texas, 28-30 September. SPE-174870-MS.

https://doi.org/10.2118/174870-MS.

Montgomery, D. C., Runger, G. C., and Hubele, N.F. 2011. *Engineering Statistics*, 5th edition. Arizona: John Wiley & Sons.

Much, M. G., and Penny, G. S. 1987. Long-Term Performance of Proppants Under Simulated Reservoir Conditions. Presented at Low Permeability Reservoirs Symposium, Denver, Colorado, 18-19 May. SPE-16415-MS.

https://doi.org/10.2118/16415-MS.

Palisch, T. T., Duenckel, R. J., Bazan, L. W. et al. 2007. Determining Realistic Fracture Conductivity and Understanding its Impact on Well Performance - Theory and Field Examples. Presented at SPE Hydraulic Fracturing Technology Conference, College Station, Texas, 29-31 January. SPE-106301-MS.

https://doi.org/10.2118/106301-MS.

Parker, M. A., and McDaniel, B. W. 1987. Fracturing Treatment Design Improved by Conductivity Measurements Under In-Situ Conditions. Presented at SPE Annual Technical Conference and Exhibition, Dallas, Texas, 27-30 September.SPE-16901-MS.

https://doi.org/10.2118/16901-MS.

Penny, G. S. 1987. An Evaluation of the Effects of Environmental Conditions and Fracturing Fluids Upon the Long-Term Conductivity of Proppants. SPE Annual Technical Conference and Exhibition, Dallas, Texas, 27-30 September. SPE-16900-MS.

https://doi.org/10.2118/16900-MS.

R Core Team 2017. R: A language and environment for statistical computing. R Foundation for S tatistical Computing, Vienna, Austria.

https://www.R-project.org/

Rivers, M., Zhu, D., and Hill, A. D. 2012. Proppant Fracture Conductivity With High Proppant L oading and High Closure Stress. Presented at SPE Hydraulic Fracturing Technology Conference, The Woodlands, Texas, 6-8 February. SPE-151972.

https://doi.org/10.2118/151972-MS.

Rubin, D.B. 1987. *Multiple Imputation for Nonresponse in Surveys.* New York: John Wiley and Sons.

Stim Lab Proppant and Fluid Consortia. 1997-2006.

U.S. Energy Information Administration. 2016. Hydraulic fracturing accounts for about half of current U.S. crude oil production. *U.S. Energy Information Administration*, https://www.eia.gov/todayinenergy/detail.php?id=25372 (accessed 15 March 2017).

Vlis, A. C., Haafkens, R., Schipper, B. A. et al. 1975. Criteria For Proppant Placement and Fracture Conductivity. Presented at Fall Meeting of the Society of Petroleum Engineers of AIME, Dallas, Texas, 28 September-1 October. SPE-5637-MS.

https://doi.org/10.2118/5637-MS.

Volk, L. J., Raible, C. J., Carroll, H. B. et al. 1981. Embedment of High Strength Proppant Into Low-Permeability Reservoir Rock. Presented at SPE/DOE Low Permeability Gas Reservoirs Symposium, Denver, Colorado, 27-29 May. SPE-9867-MS.

https://doi.org/10.2118/9867-MS.

Zhang, J. 2014. Creation and Impairment of Hydraulic Fracture Conductivity in Shale Formations. Texas A & M University, College Station, TX. (August, 2014).

http://hdl.handle.net/1969.1/153365.

APPENDIX A

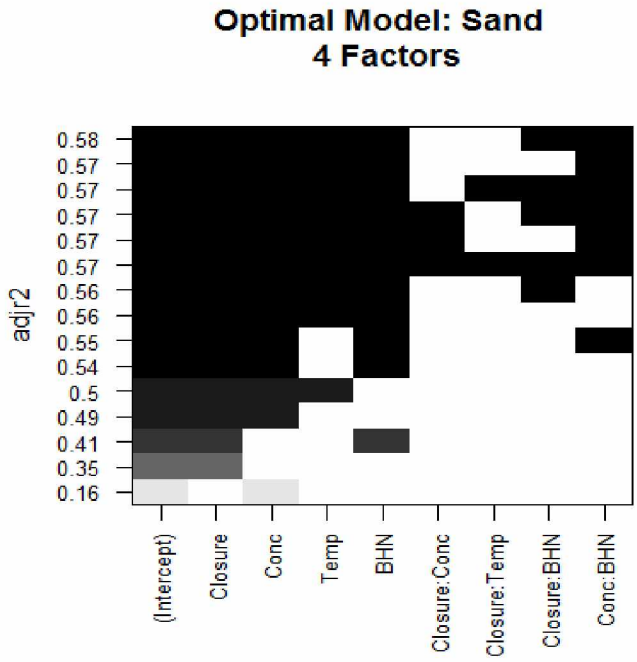

Figure A-1: Optimal model from all subsets regression for 2nd four factor model of Proppant Type-1

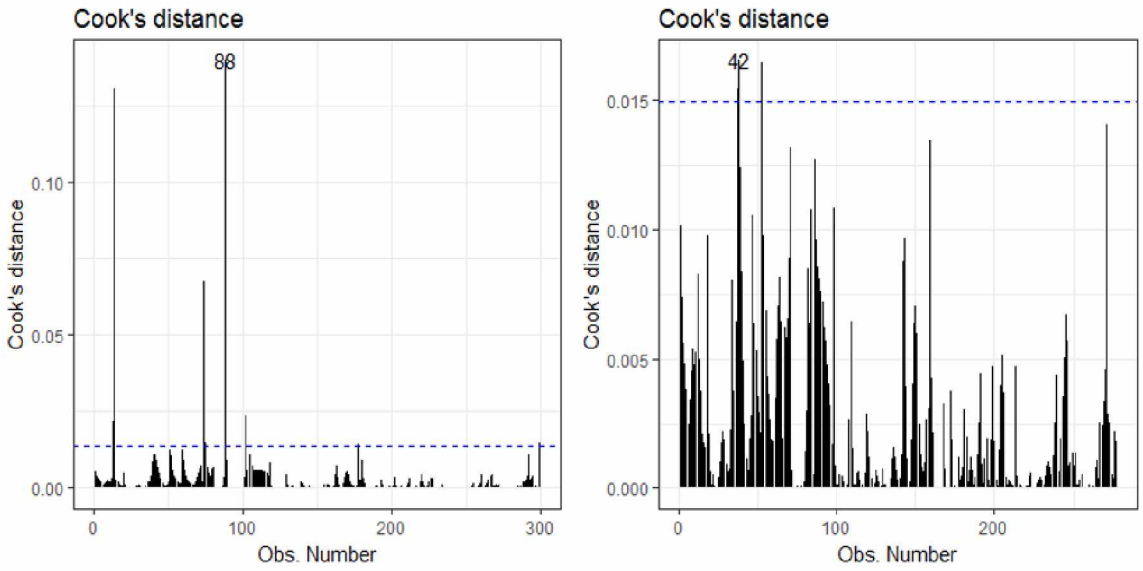

Figure A-2: Cook's distance plot for 2nd four factor model of Proppant Type-1 (Left-before deletion, Right-after deletion)

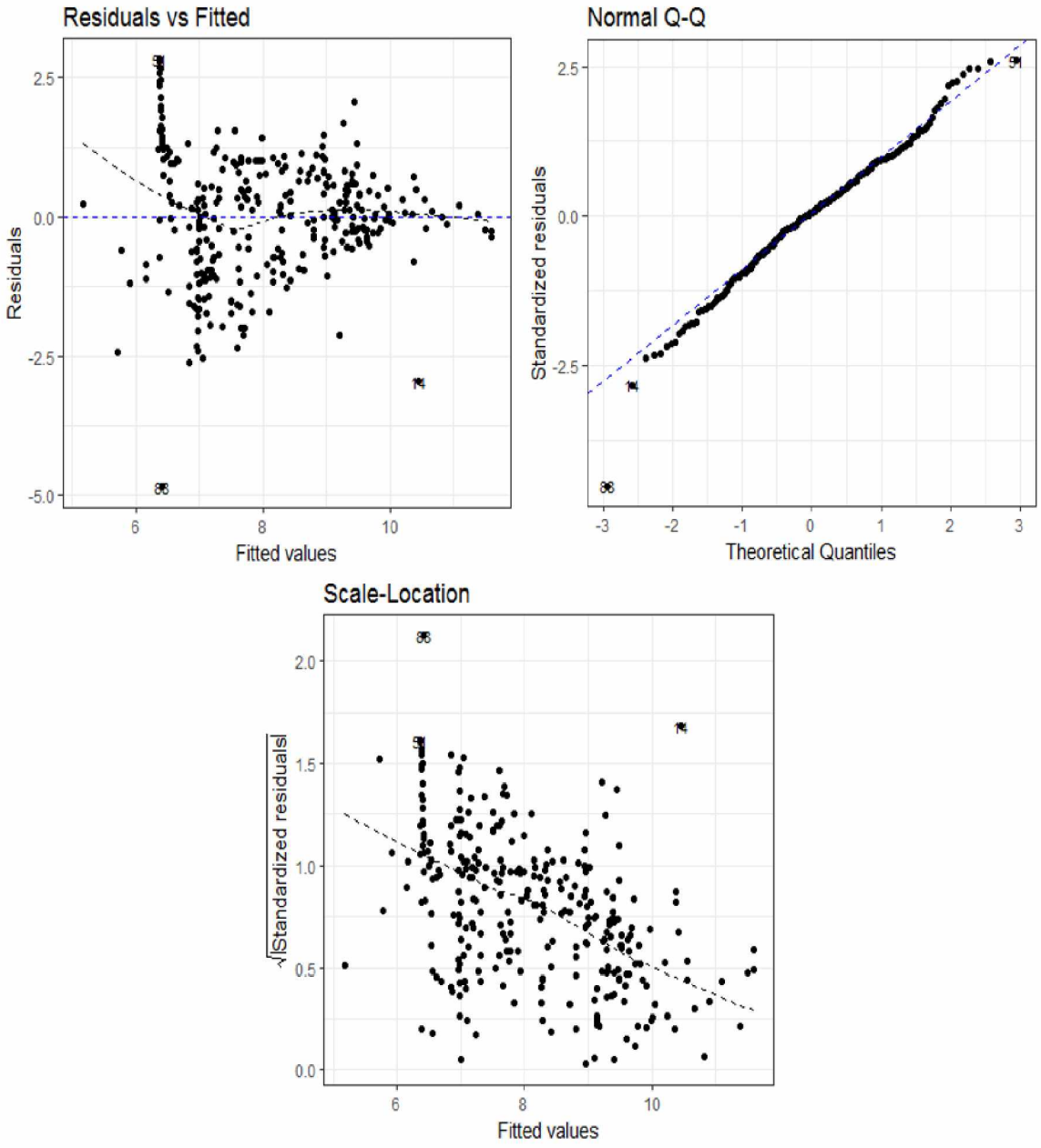

Figure A-3: Regression diagnostic before deleting influential observations as the validation of model assumptions for 2nd four factor model of Proppant Type-1, (Upper left-Linearity, Upper right-Normality, Bottom-Homoscedasticity)

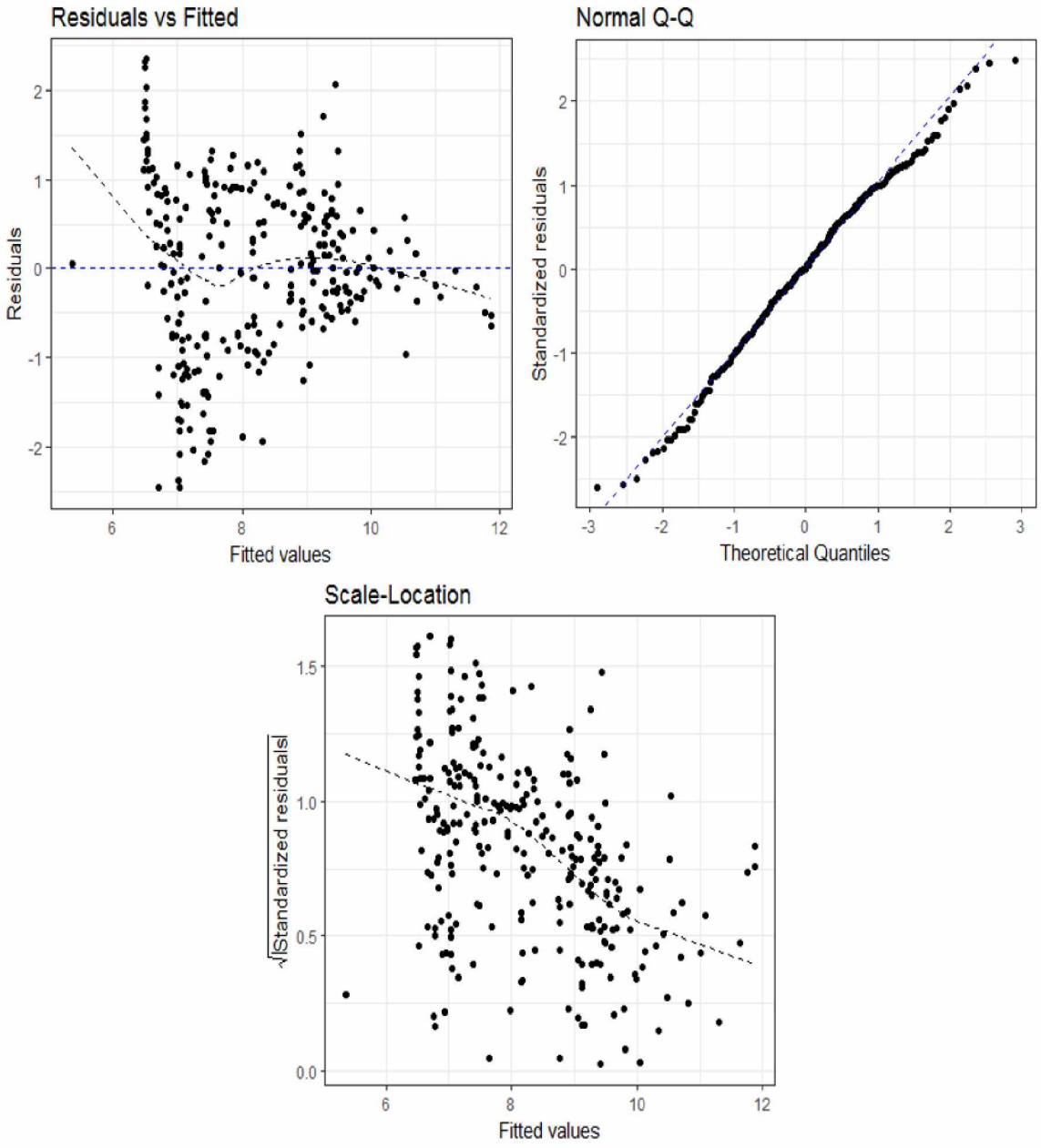

Figure A-4: Regression diagnostic after deleting influential observations as the validation of model assumptions for 2nd four factor model of Proppant Type-1, (Upper left-Linearity, Upper right-Normality, Bottom-Homoscedasticity)

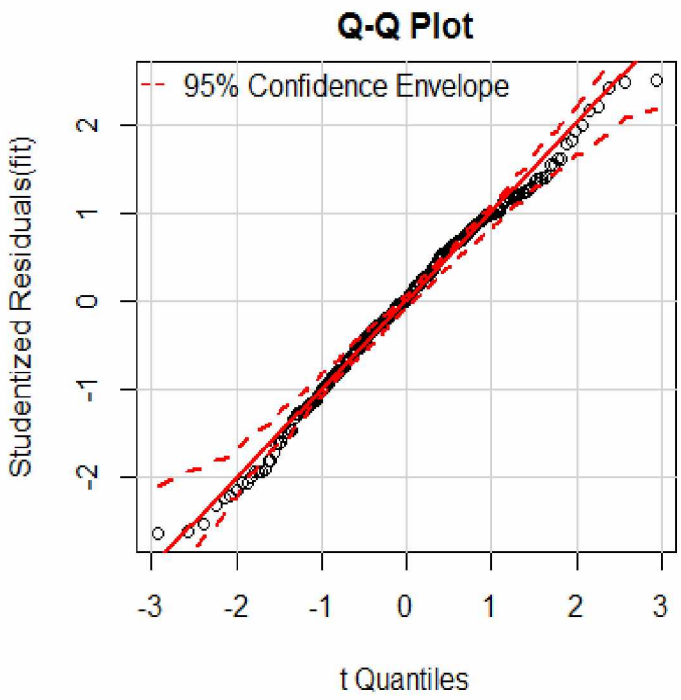

Figure A-5: Q-Q plot of studentized residuals for 2nd four factor model of Proppant Type-1

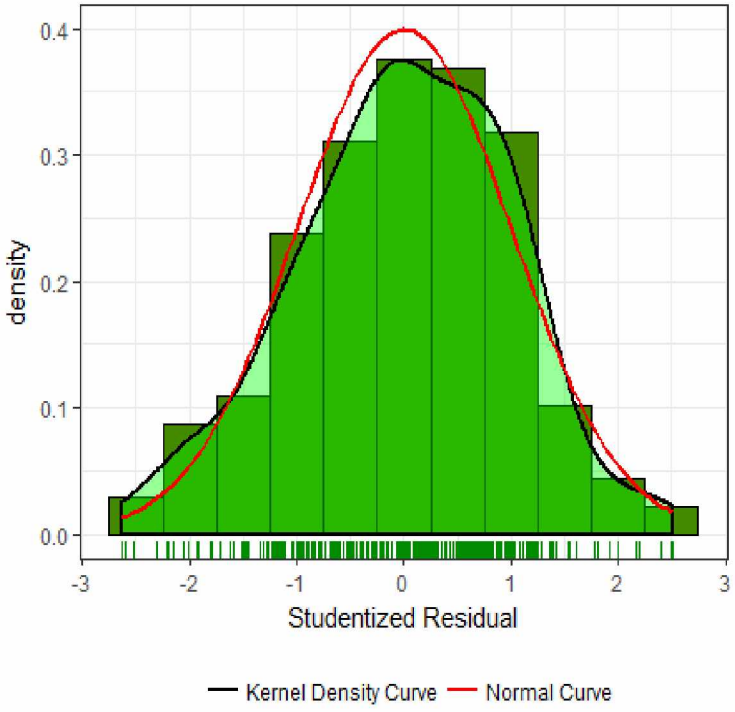

Figure A-6: Distribution of errors for 2nd four factor model of Proppant Type-1

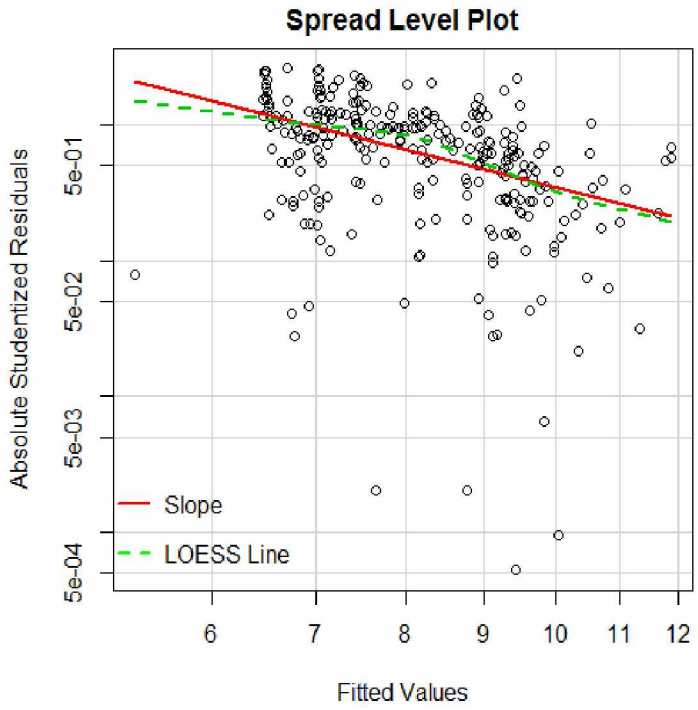

Figure A-7: Spread level plot for 2nd four factor model of Proppant Type-1

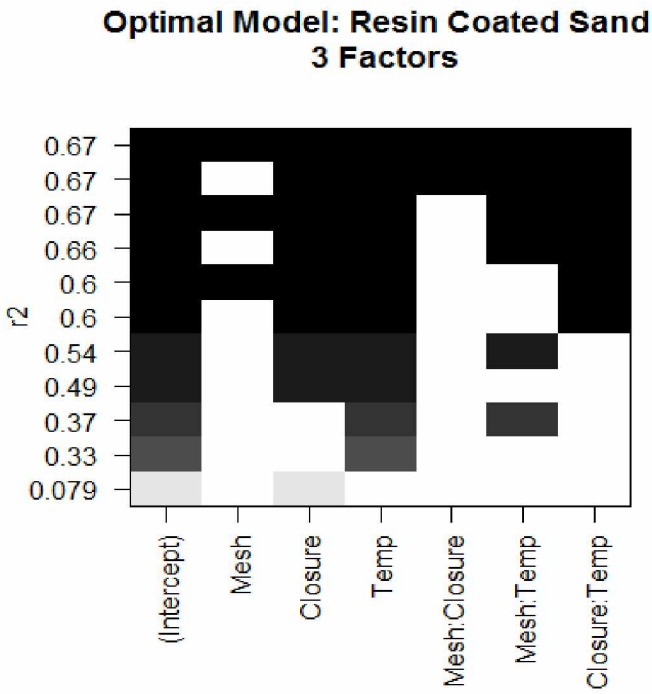

Figure A-8: Optimal model from all subsets regression for three factor model of Proppant Type-2

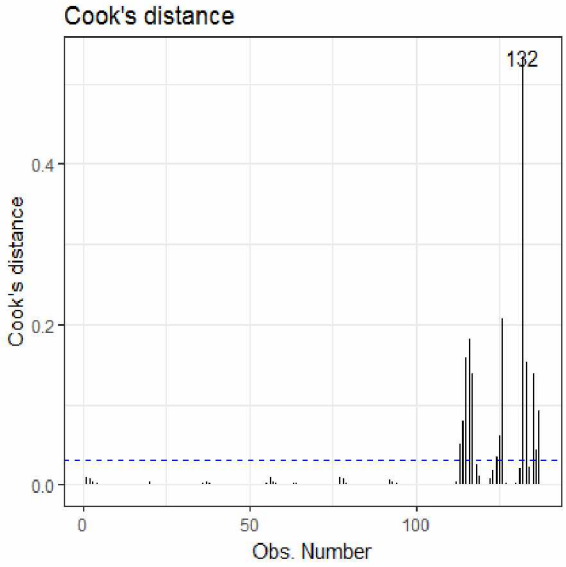

Figure A-9: Cook's distance plot for three factor model of Proppant Type-2

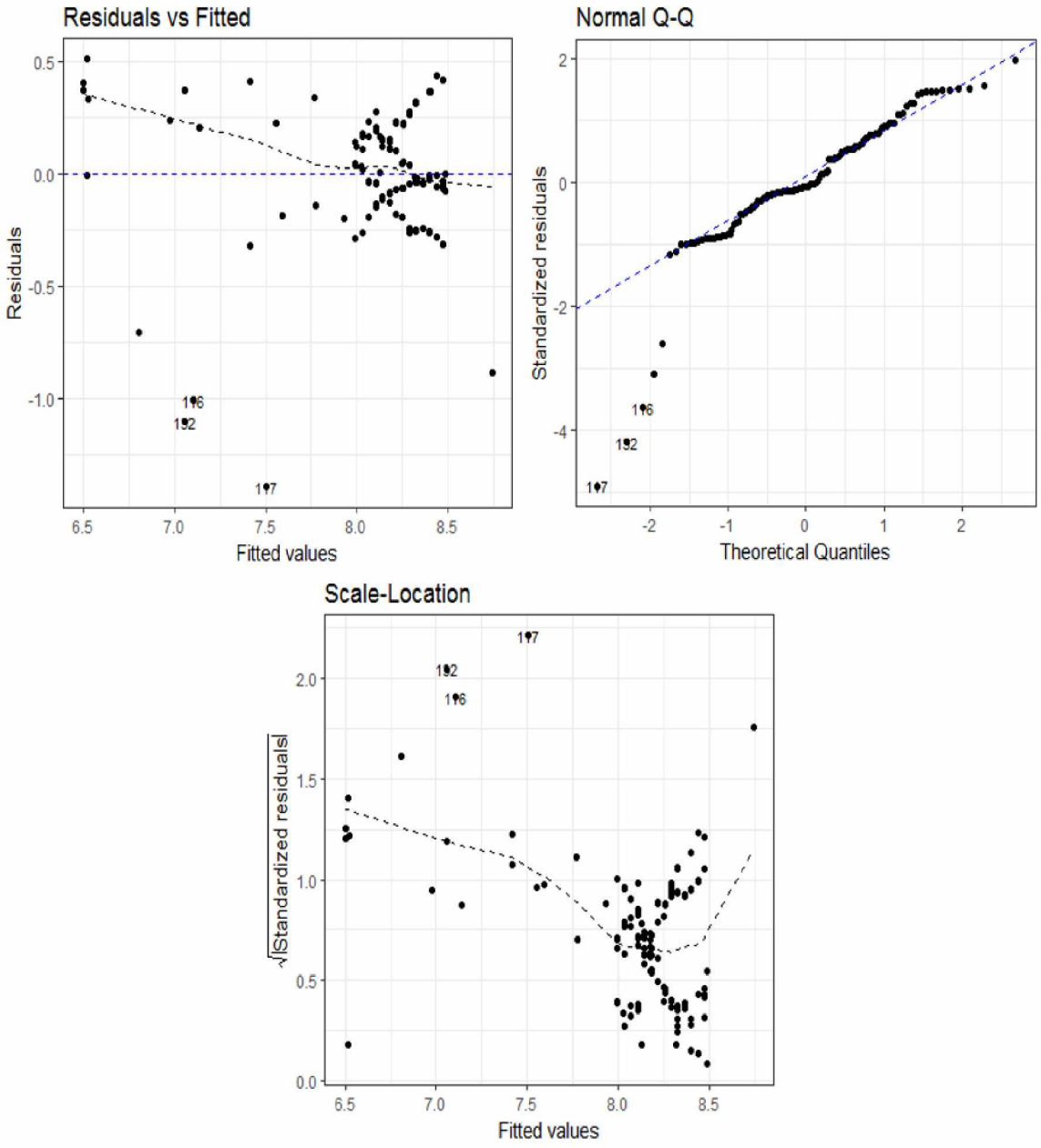

Figure A-10: Regression diagnostic as the validation of model assumptions for three factor model of Proppant Type-2, (Upper left-Linearity, Upper right-Normality, Bottom-Homoscedasticity)

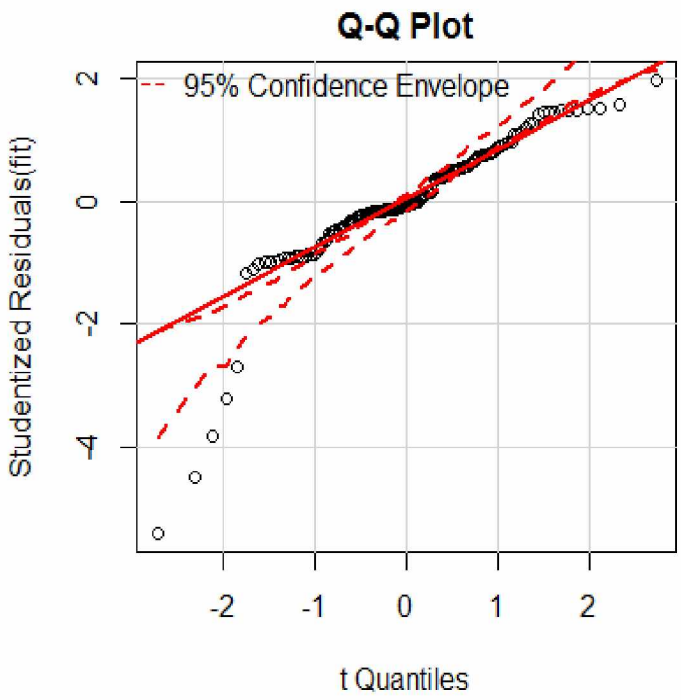

Figure A-11: Q-Q plot of studentized residuals for three factor model of Proppant Type-2

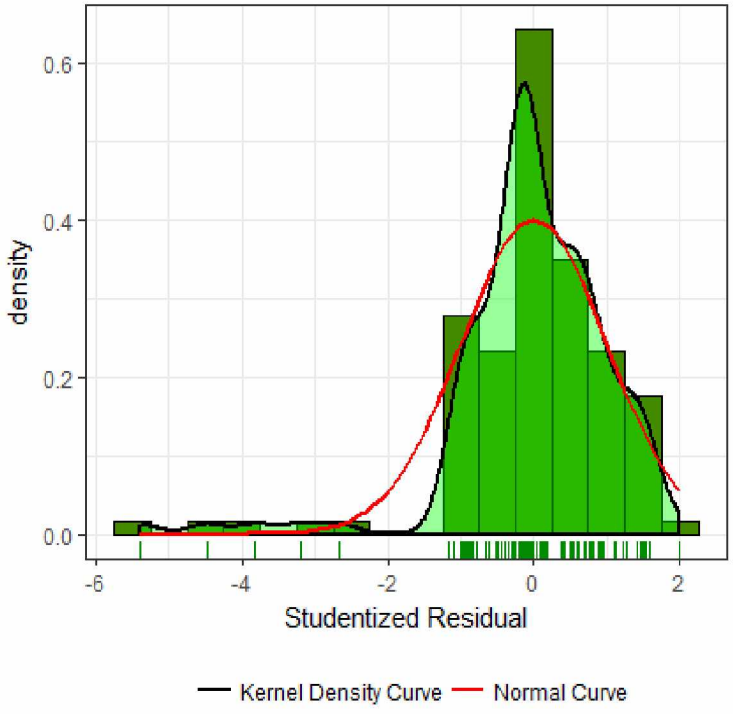

Figure A-12: Distribution of errors for three factor model of Proppant Type-2

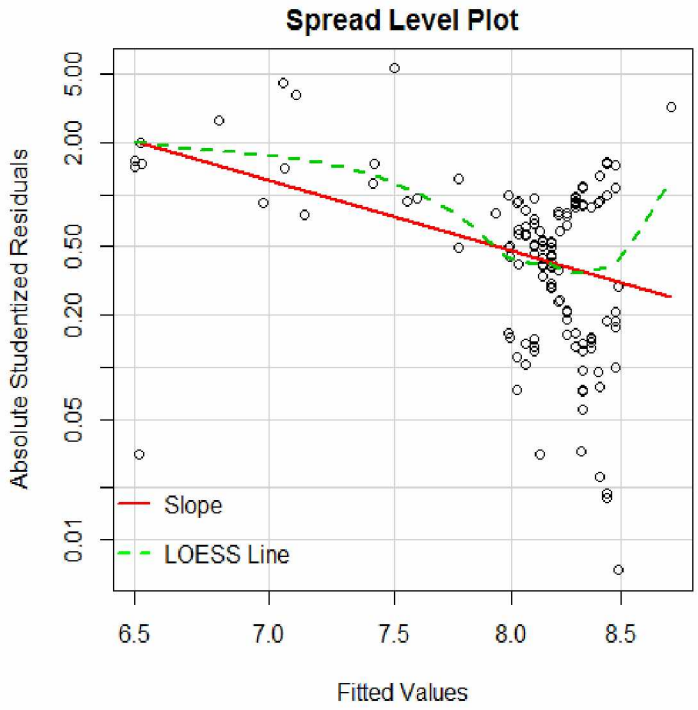

Figure A-13: Spread level plot for three factor model of Proppant Type-2

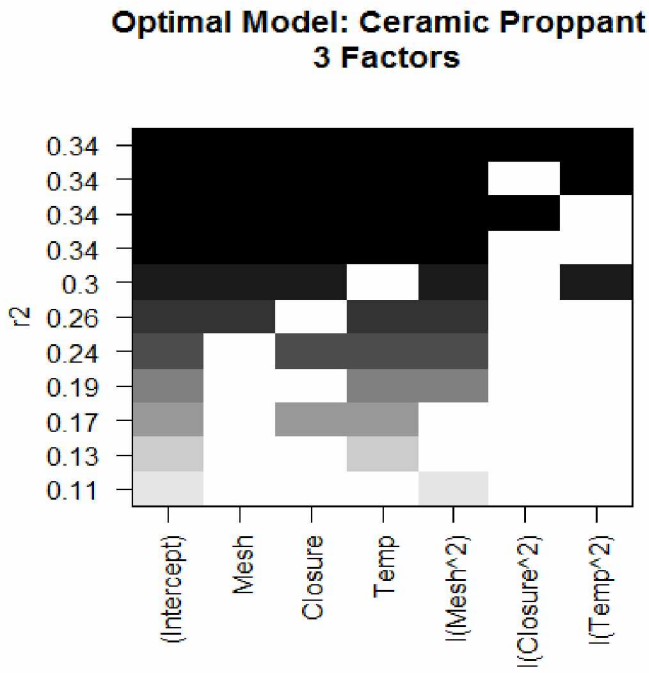

Figure A-14: Optimal model from all subsets regression for three factor model of Proppant Type-3

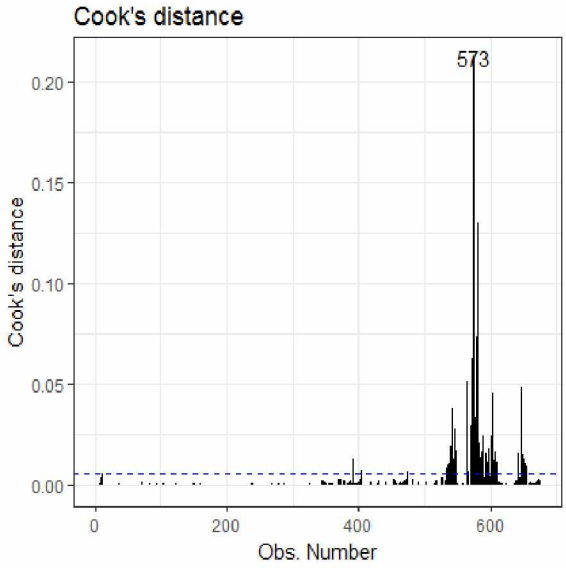

Figure A-15: Cook's distance plot for three factor model of Proppant Type-3

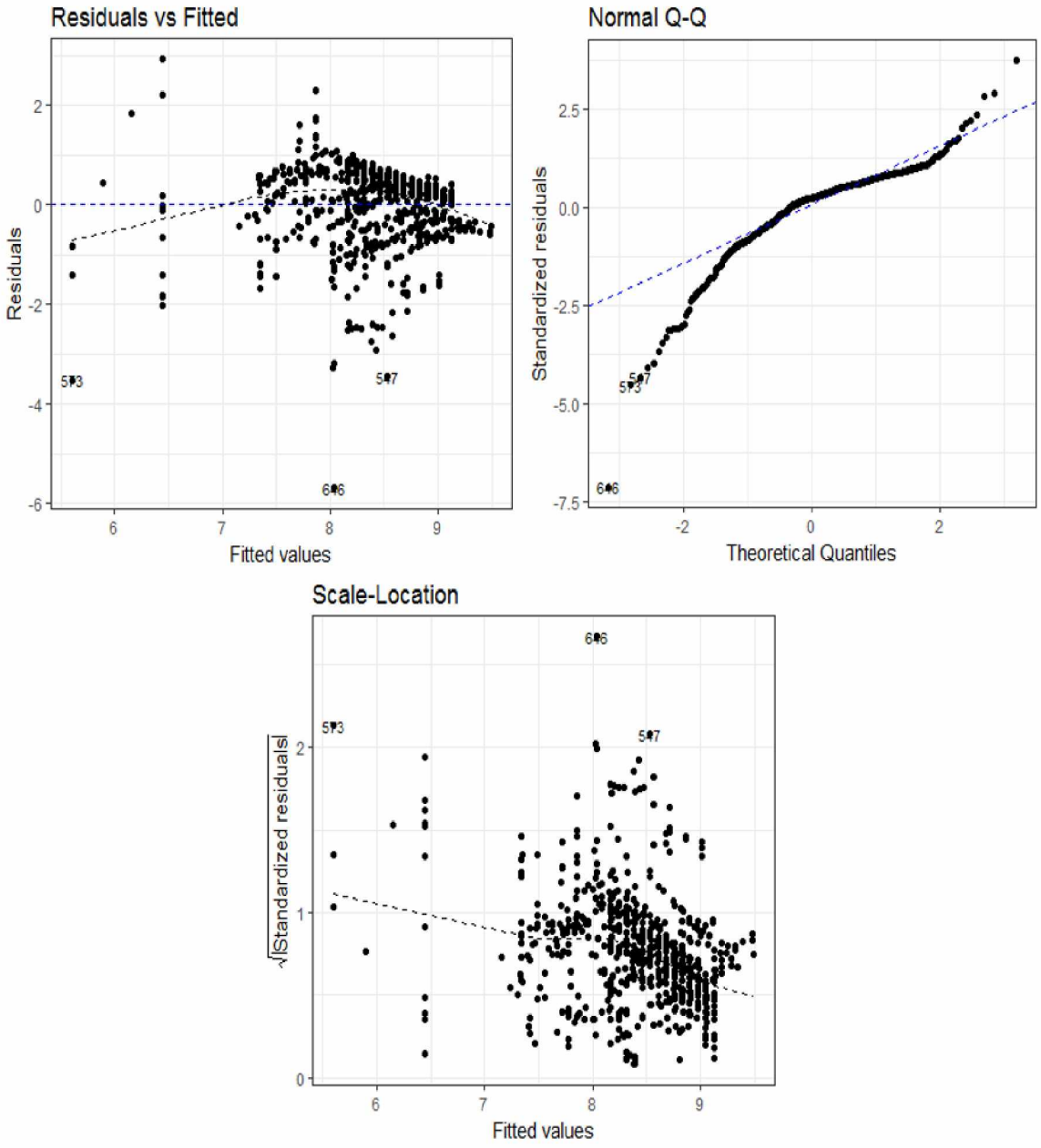

Figure A-16: Regression diagnostic as the validation of model assumptions for three factor model of Proppant Type-3, (Upper left-Linearity, Upper right-Normality, Bottom-Homoscedasticity)

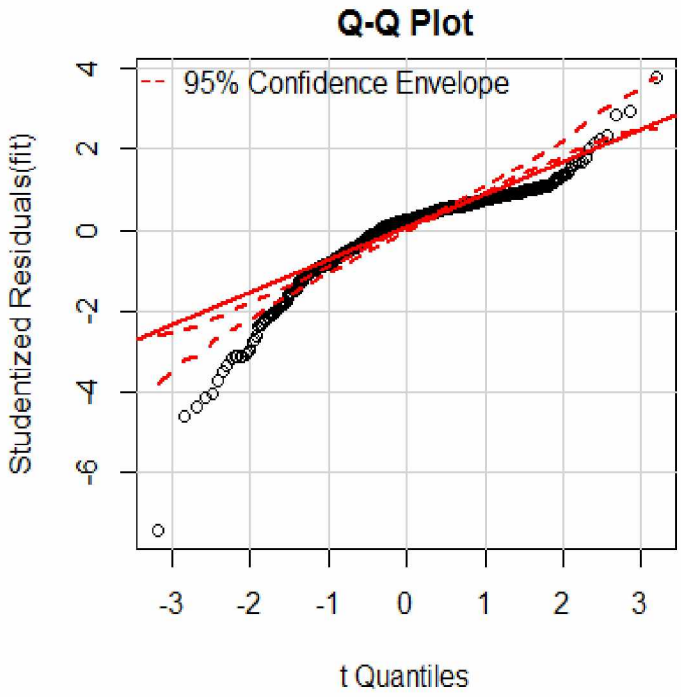

Figure A-17: Q-Q plot of studentized residuals for three factor model of Proppant Type-3

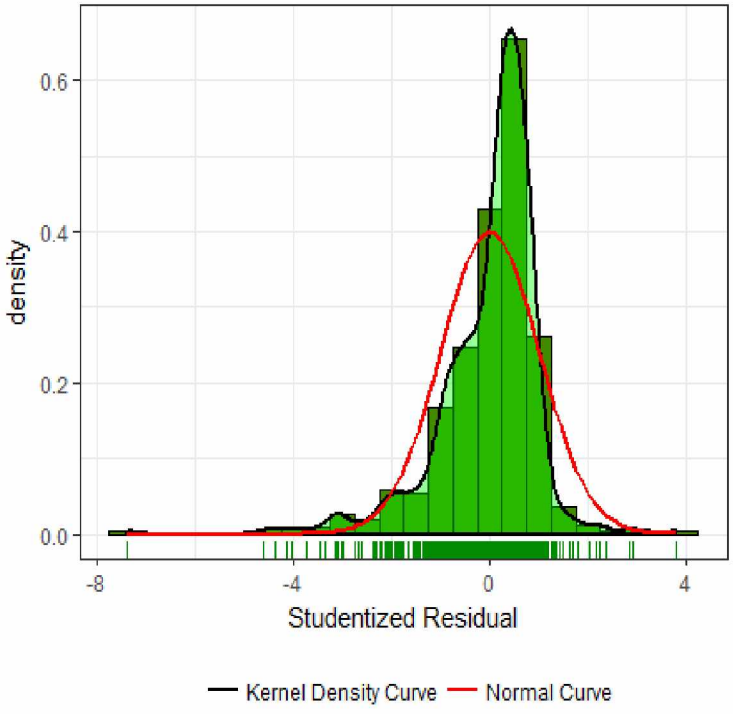

Figure A-18: Distribution of errors for three factor model of Proppant Type-3

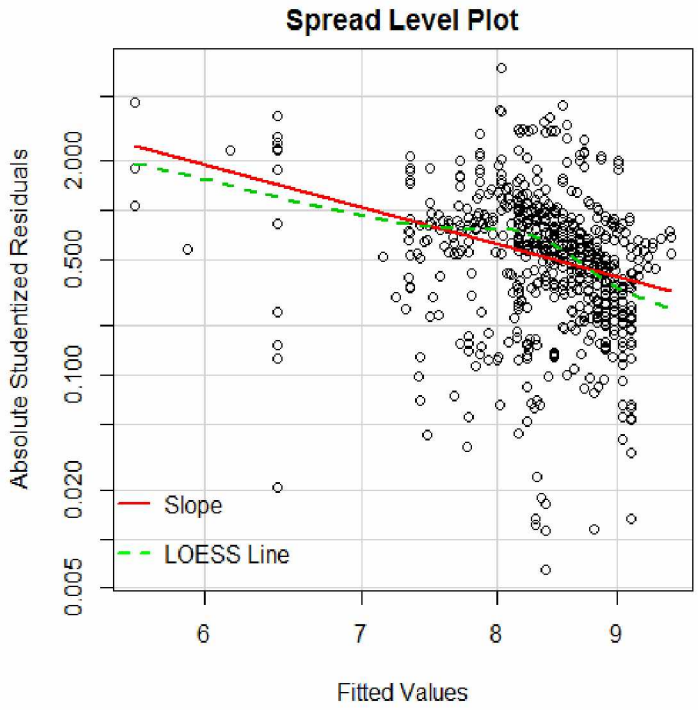

Figure A-19: Spread level plot for three factor model of Proppant Type-3

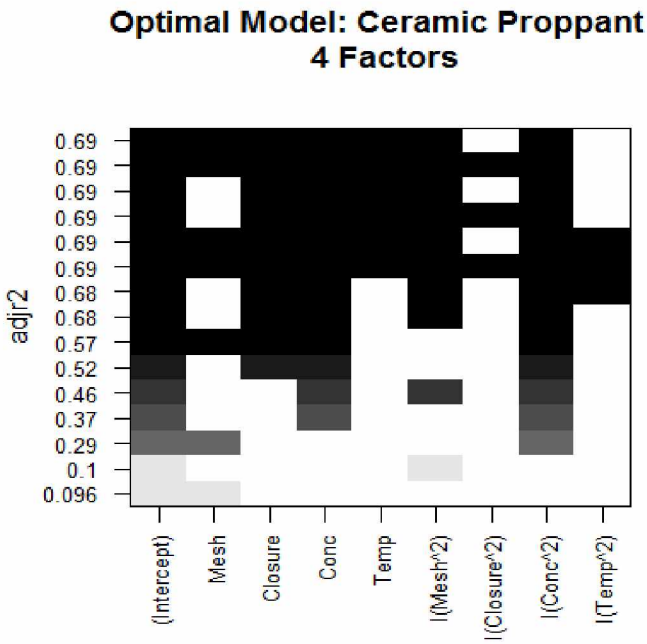

Figure A-20: Optimal model from all subsets regression for four factor model of Proppant Type-3

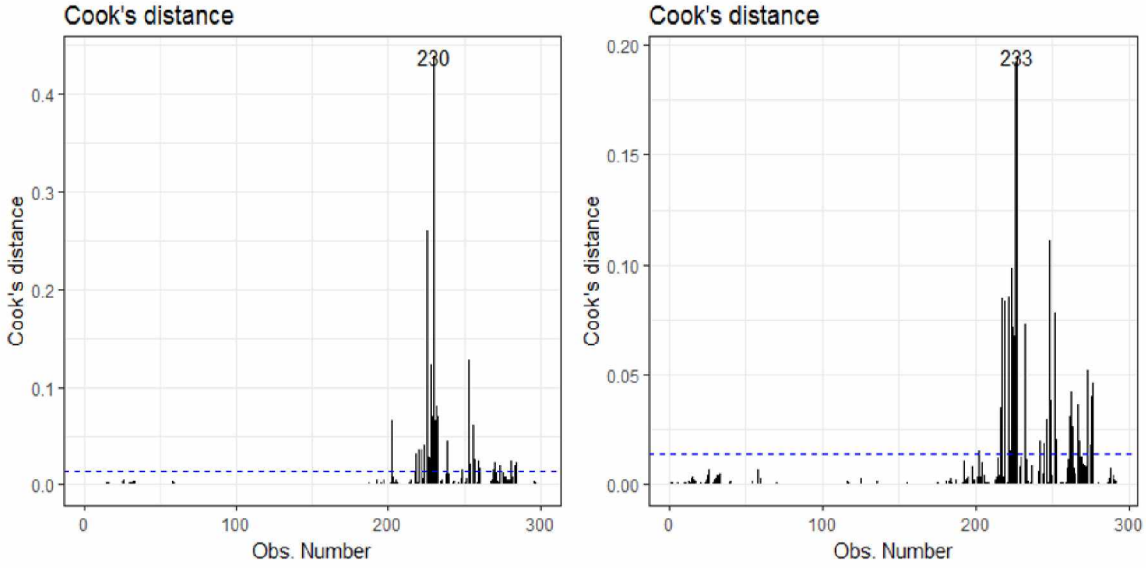

Figure A-21: Cook's distance plot for four factor model of Proppant Type-3 (Left-before deletion, Right-after deletion)

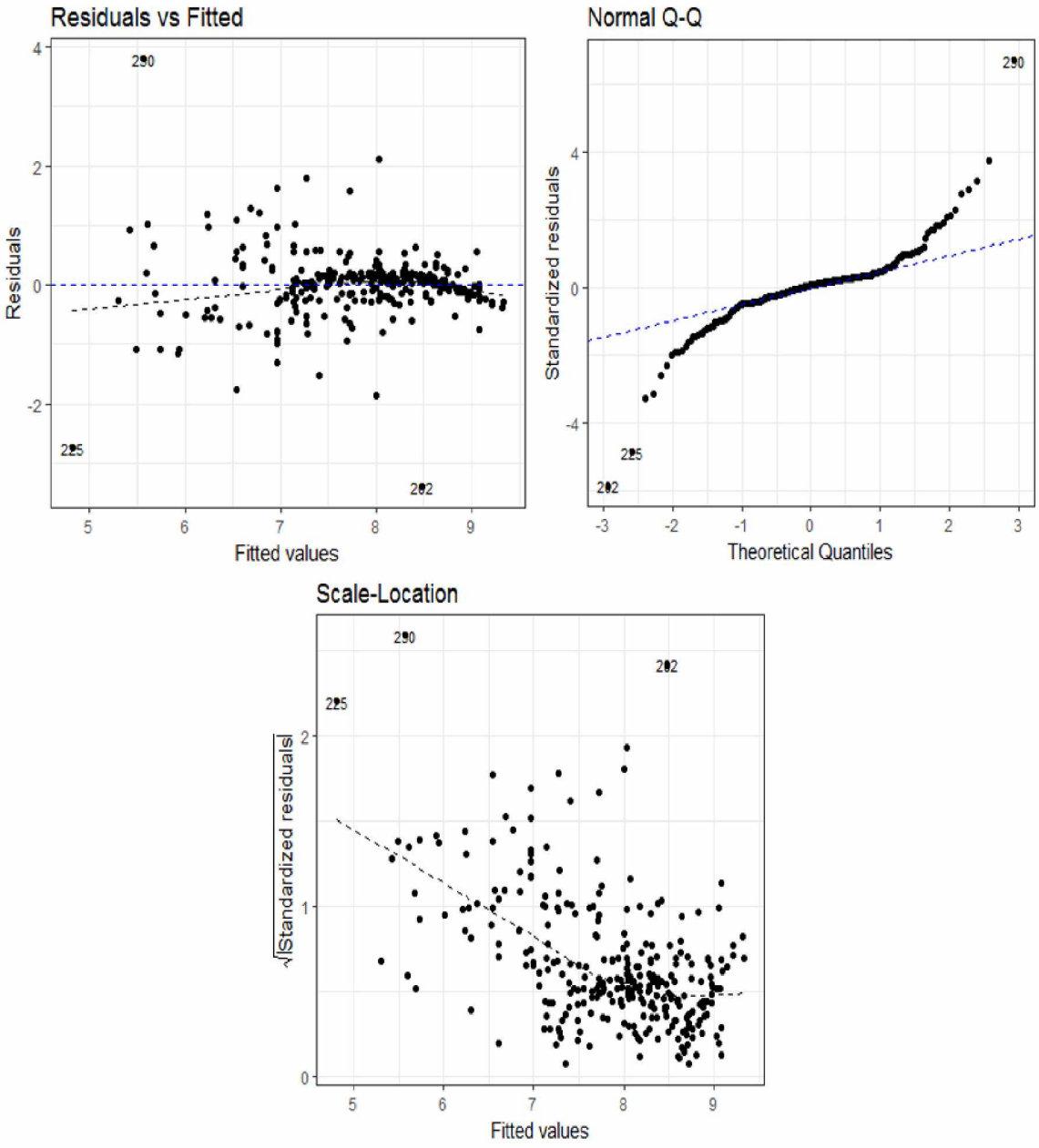

Figure A-22: Regression diagnostic before deleting influential observations as the validation of model assumptions for four factor model of Proppant Type-3, (Upper left-Linearity, Upper right-Normality, Bottom-Homoscedasticity)

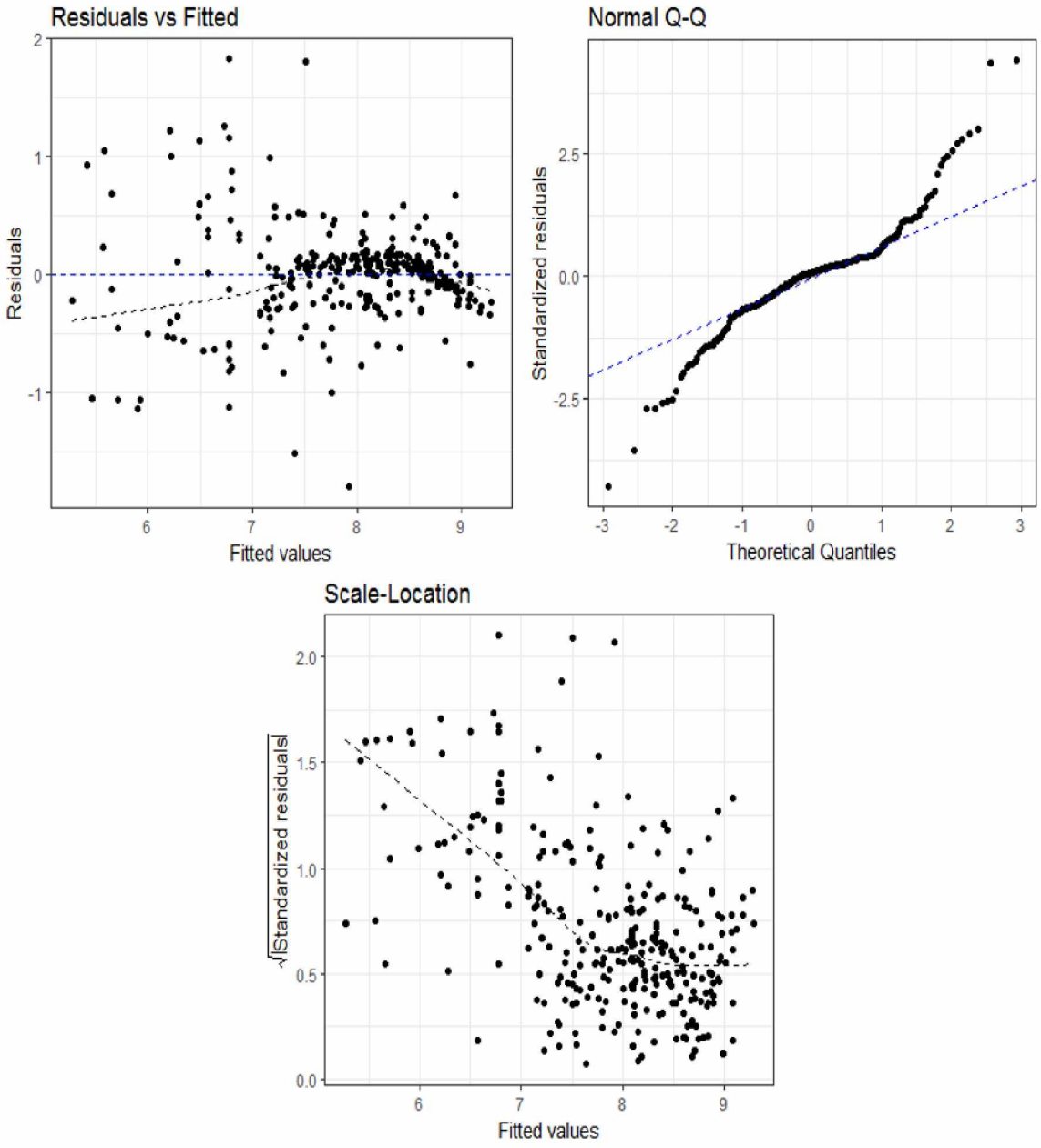

Figure A-23: Regression diagnostic after deleting influential observations as the validation of model assumptions for four factor model of Proppant Type-3, (Upper left-Linearity, Upper right-Normality, Bottom-Homoscedasticity)

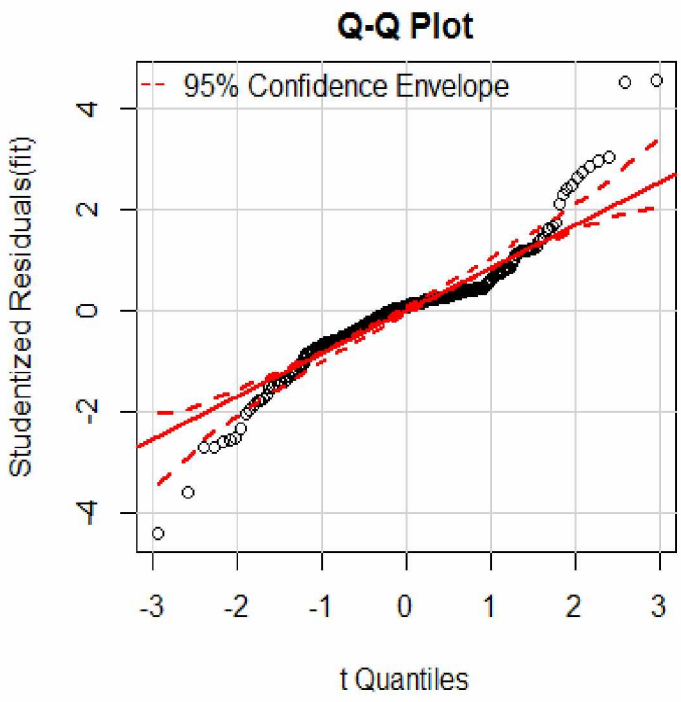

Figure A-24: Q-Q plot of studentized residuals for four factor model of Proppant Type-3

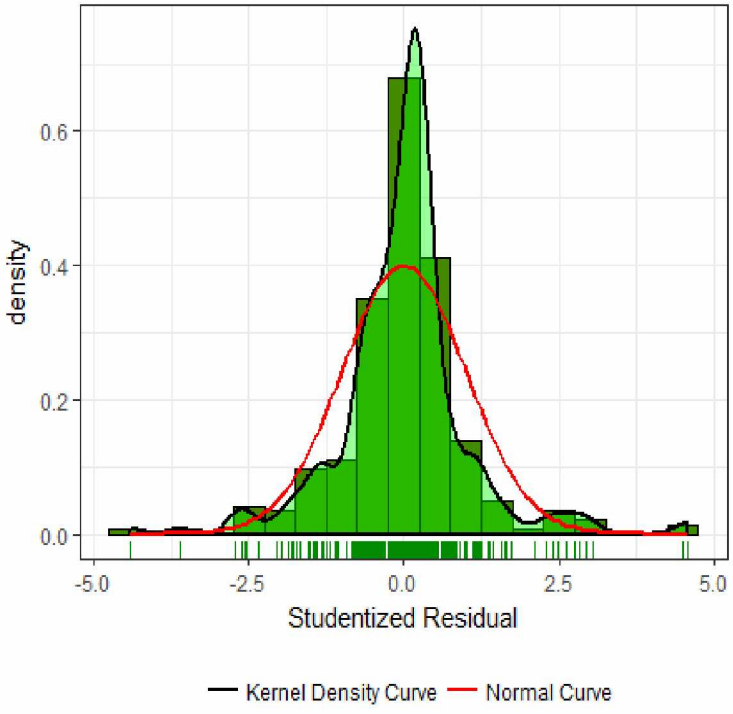

Figure A-25: Distribution of errors for four factor model of Proppant Type-3

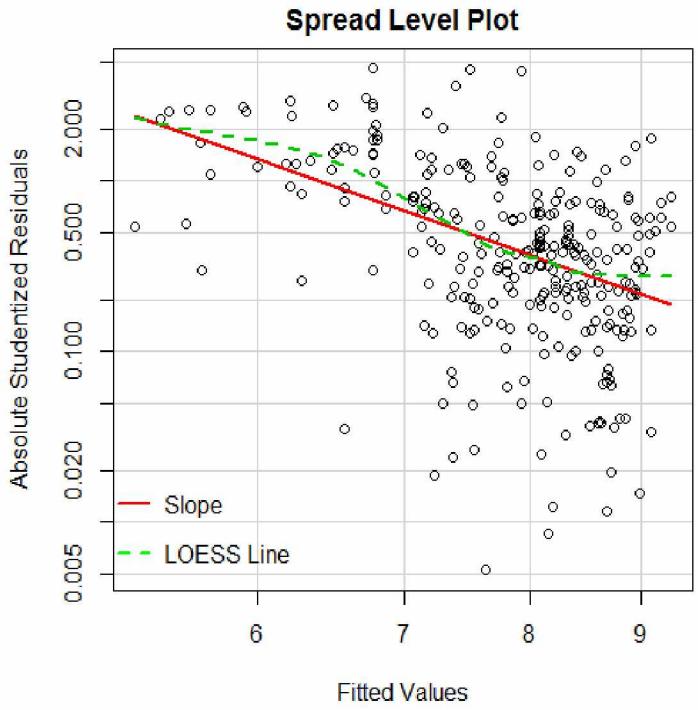

Figure A-26: Spread level plot for four factor model of Proppant Type-3

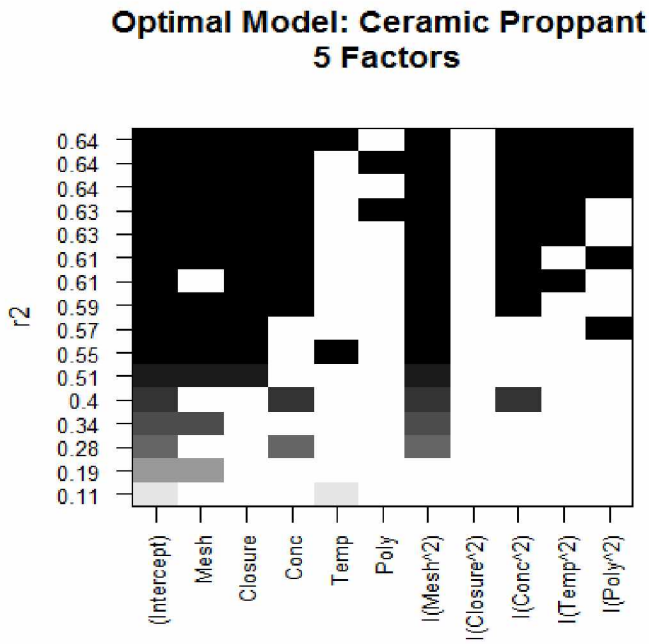

Figure A-27: Optimal model from all subsets regression for five factor model of Proppant Type-3

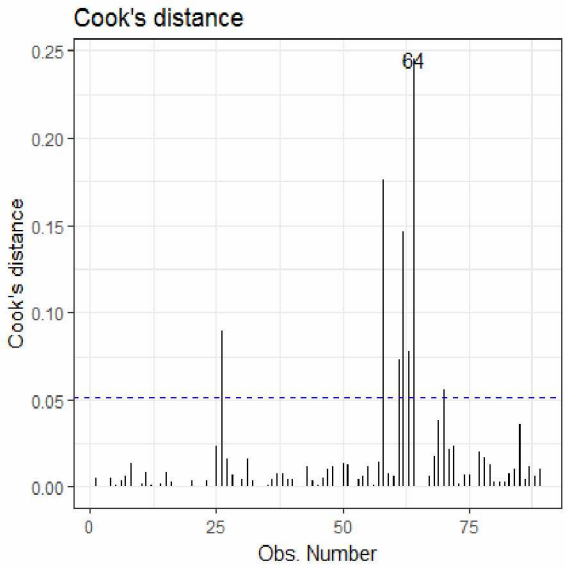

Figure A-28: Cook's distance plot for five factor model of Proppant Type-3

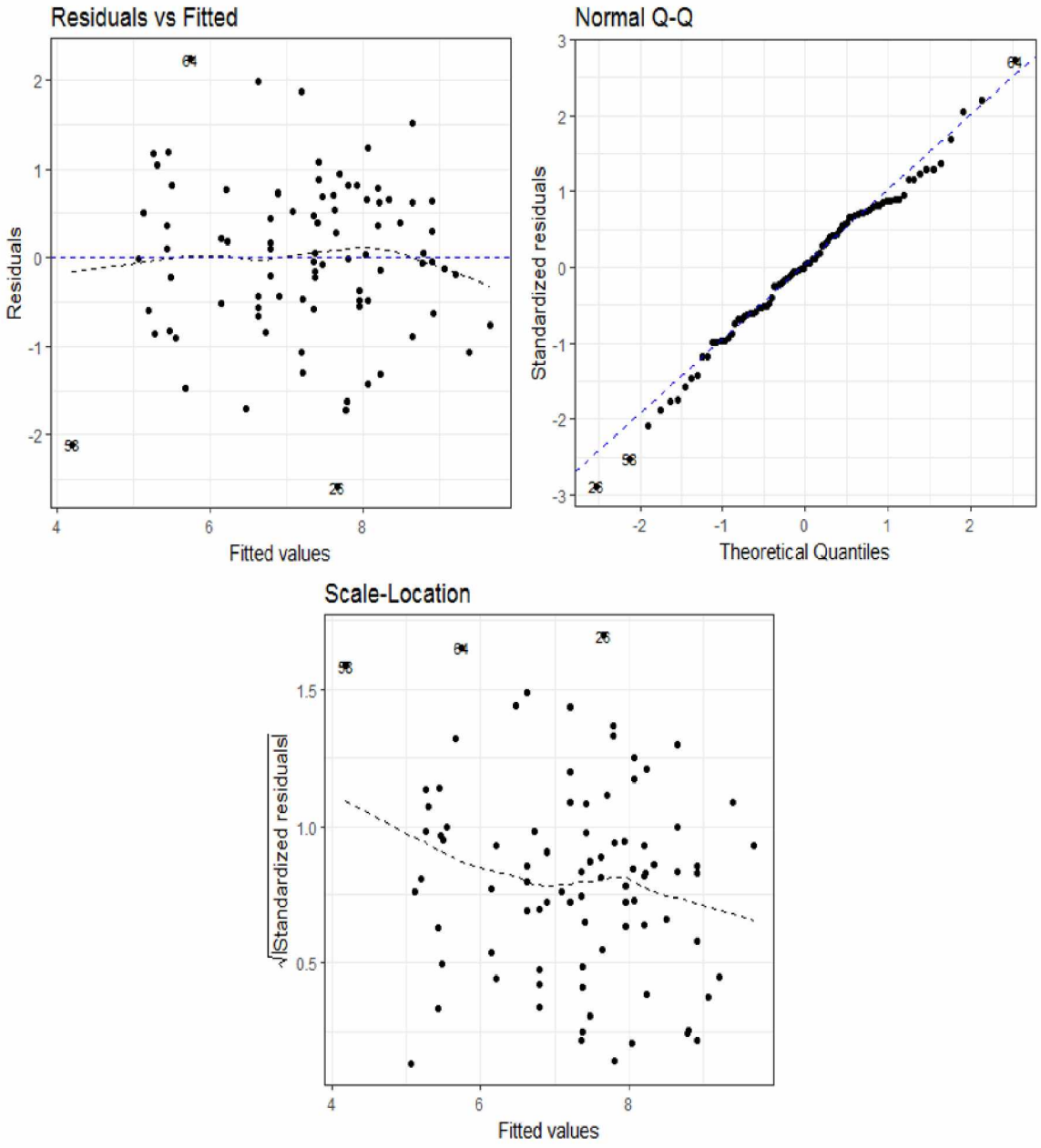

Figure A-29: Regression diagnostic as the validation of model assumptions for five factor model of Proppant Type-3, (Upper left-Linearity, Upper right-Normality, Bottom-Homoscedasticity)

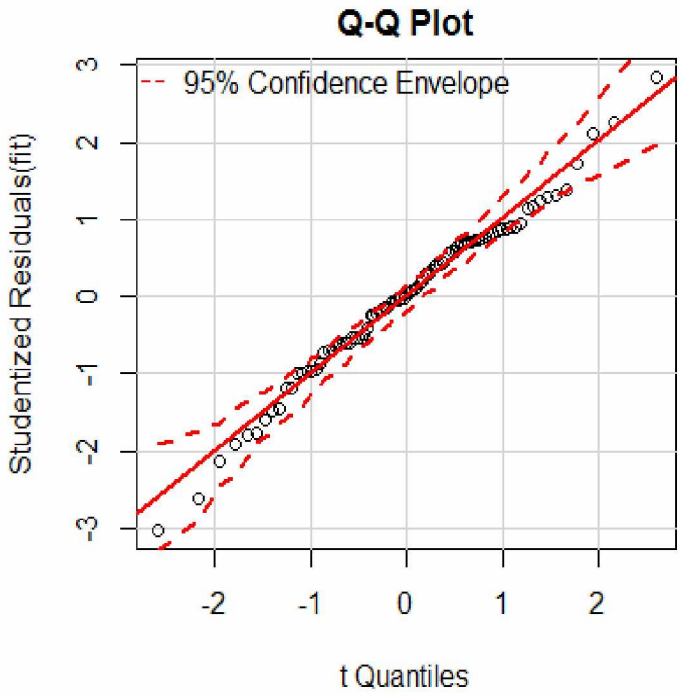

Figure A-30: Q-Q plot of studentized residuals for five factor model of Proppant Type-3

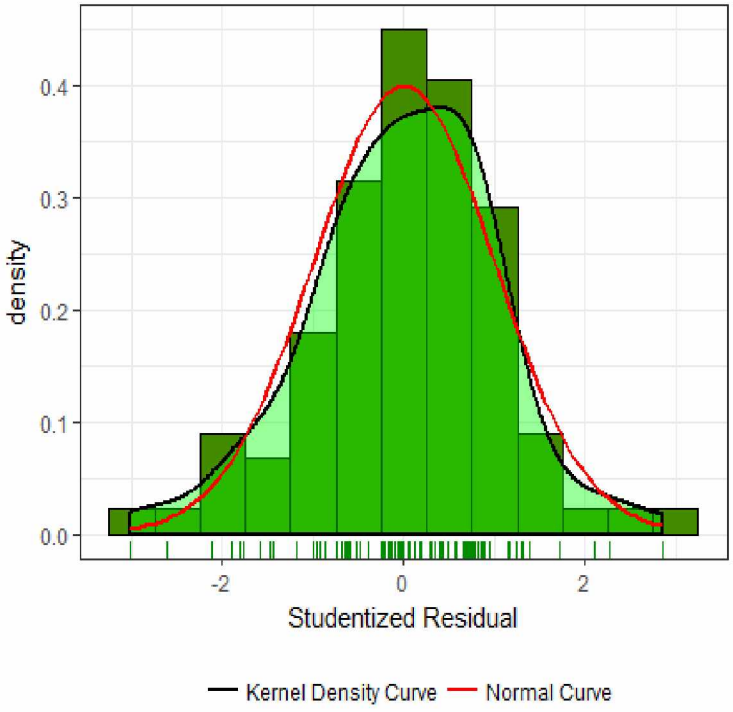

Figure A-31: Distribution of errors for five factor model of Proppant Type-3

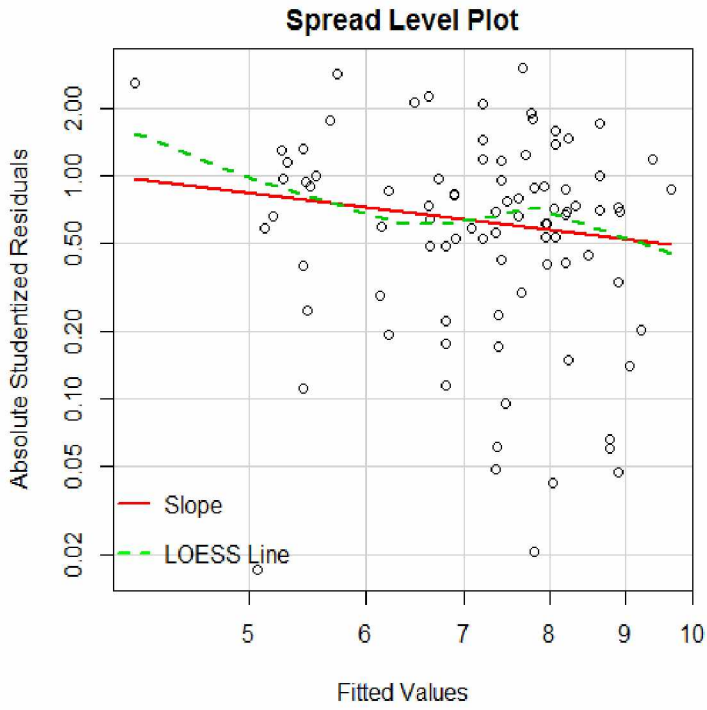

Figure A-32: Spread level plot for five factor model of Proppant Type-3

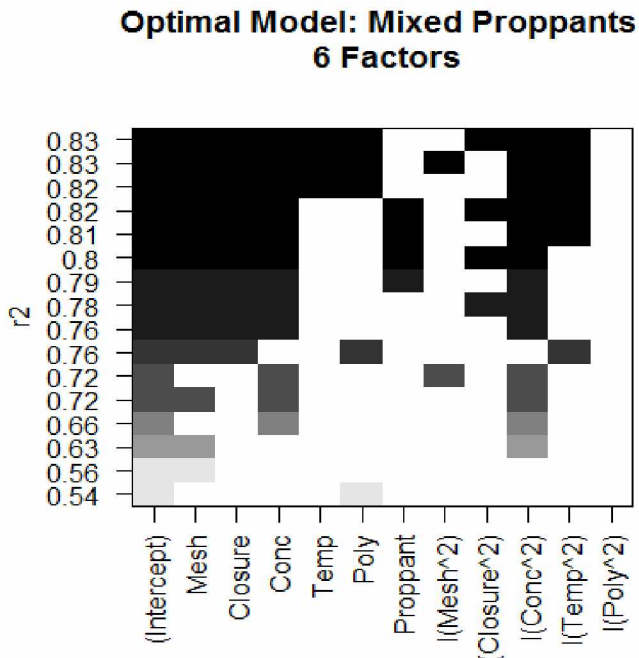

Figure A-33: Optimal model from all subsets regression for six factor model of All Proppant Type

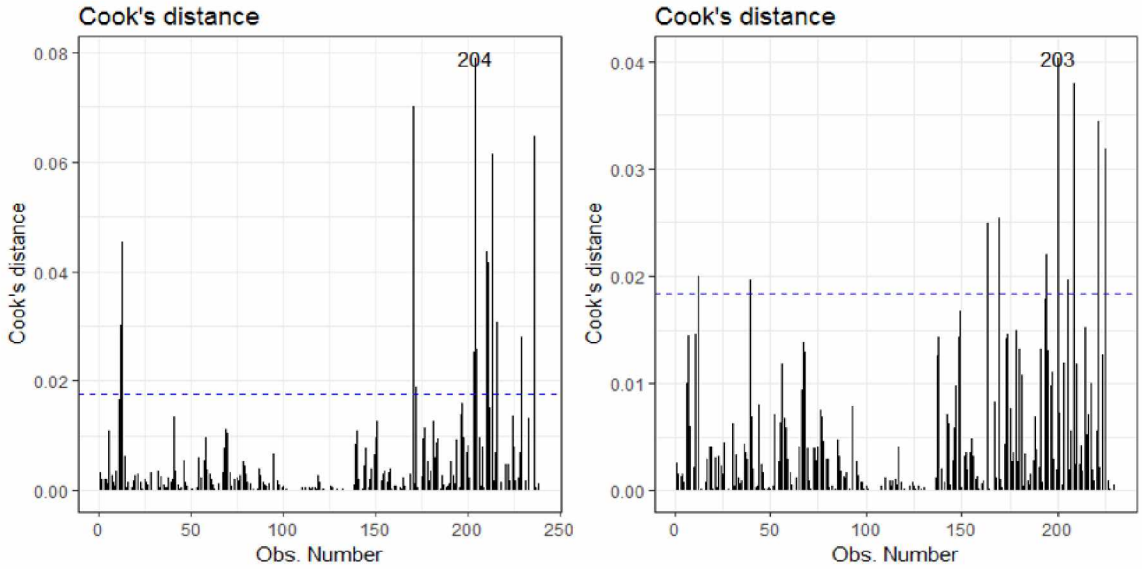

Figure A-34: Cook's distance plot for six factor model of All Proppant Type (Left-before deletion, Right-after deletion)

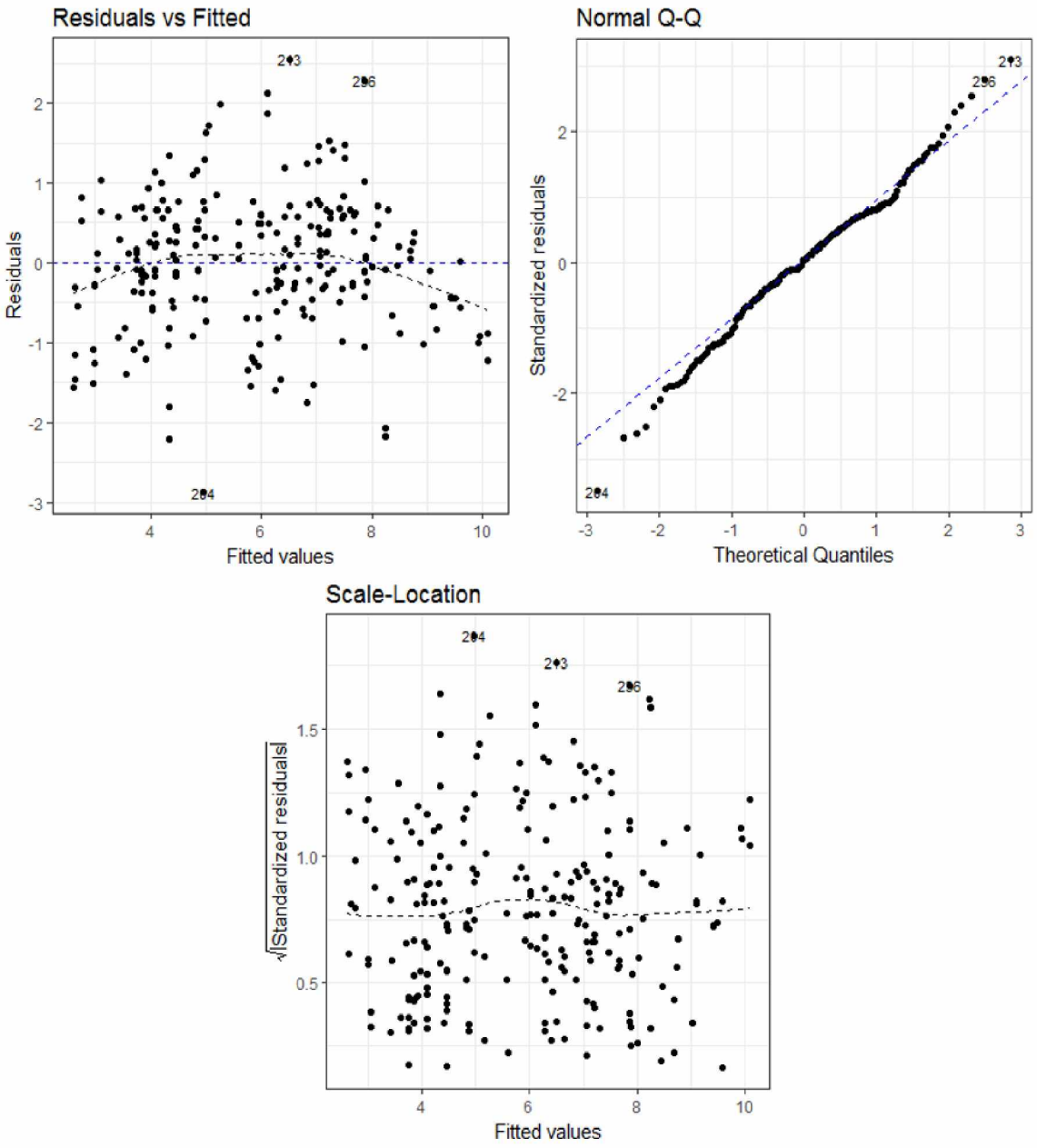

Figure A-35: Regression diagnostic before deleting influential observations as the validation of model assumptions for six factor model of All Proppant Type (Upper left-Linearity, Upper right-Normality, Bottom-Homoscedasticity)

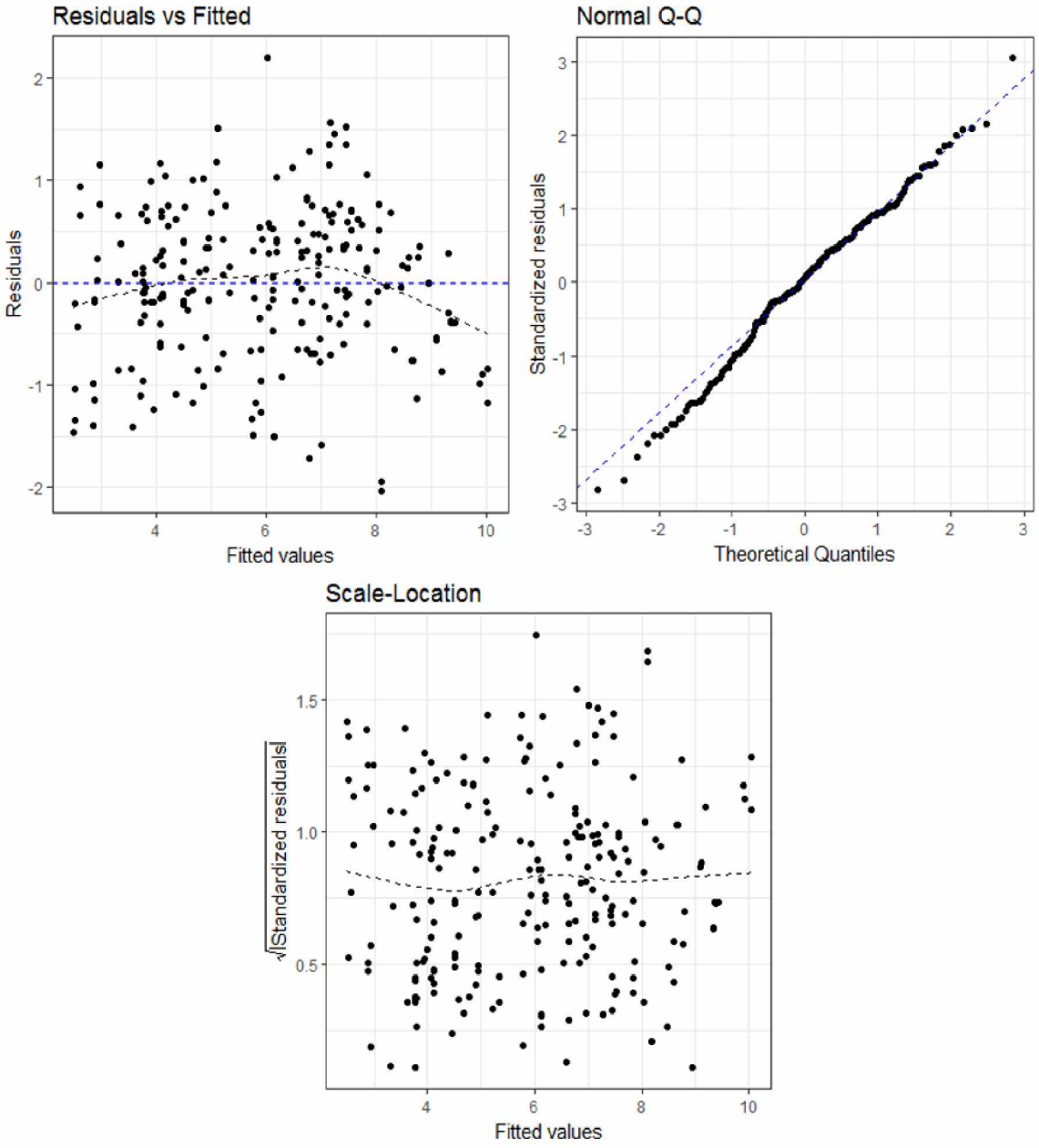

Figure A-36: Regression diagnostic after deleting influential observations as the validation of model assumptions for six factor model of All Proppant Type (Upper left-Linearity, Upper right-Normality, Bottom-Homoscedasticity)

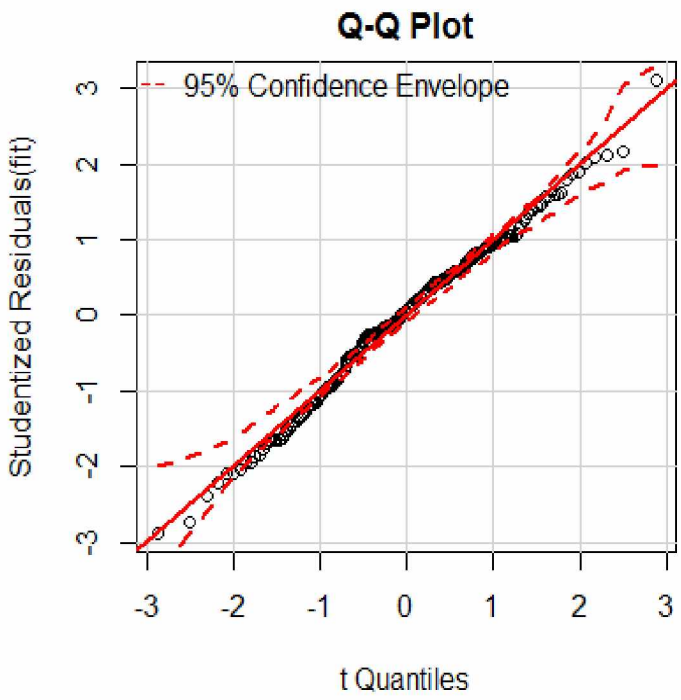

Figure A-37: Q-Q plot of studentized residuals for six factor model of All Proppant Type

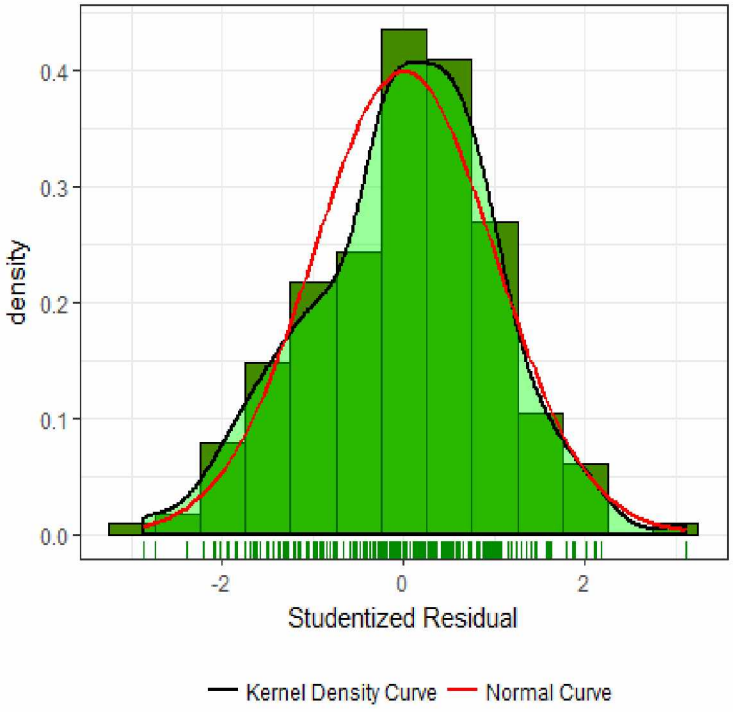

Figure A-38: Distribution of errors for six factor model of All Proppant Type

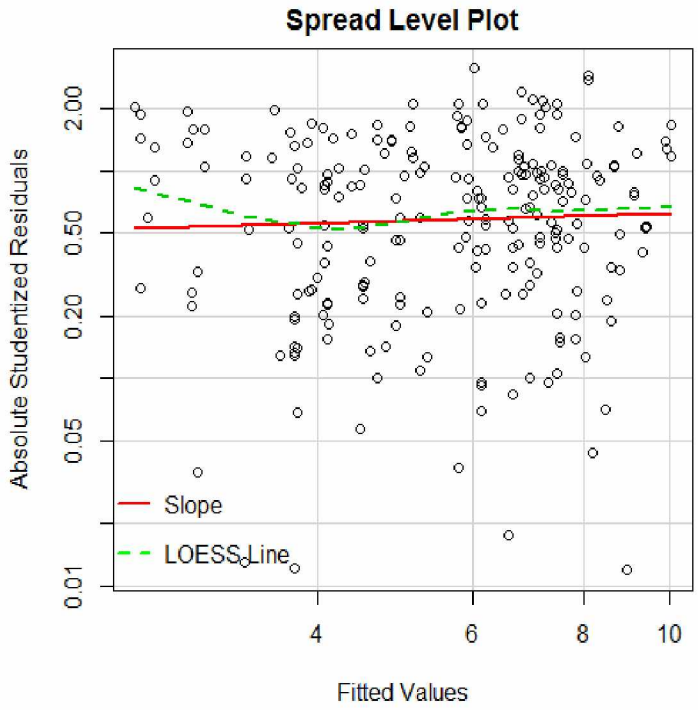

Figure A-39: Spread level plot for six factor model of All Proppant Type

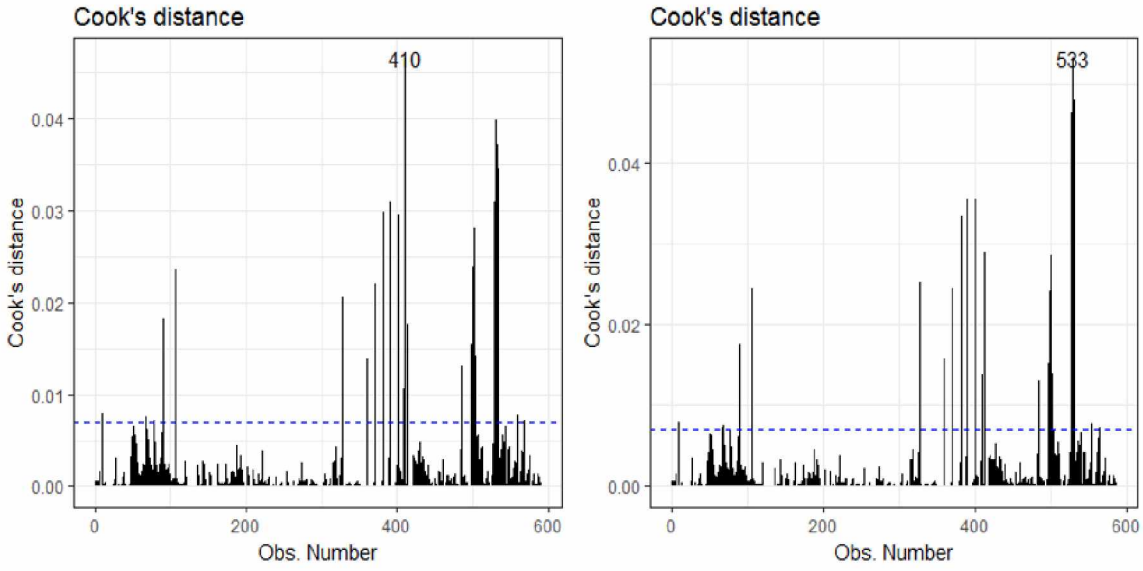

Figure A-40: Cook's distance plot for categorical imputation model of Proppant Type-1 (Left-before deletion, Right-after deletion)

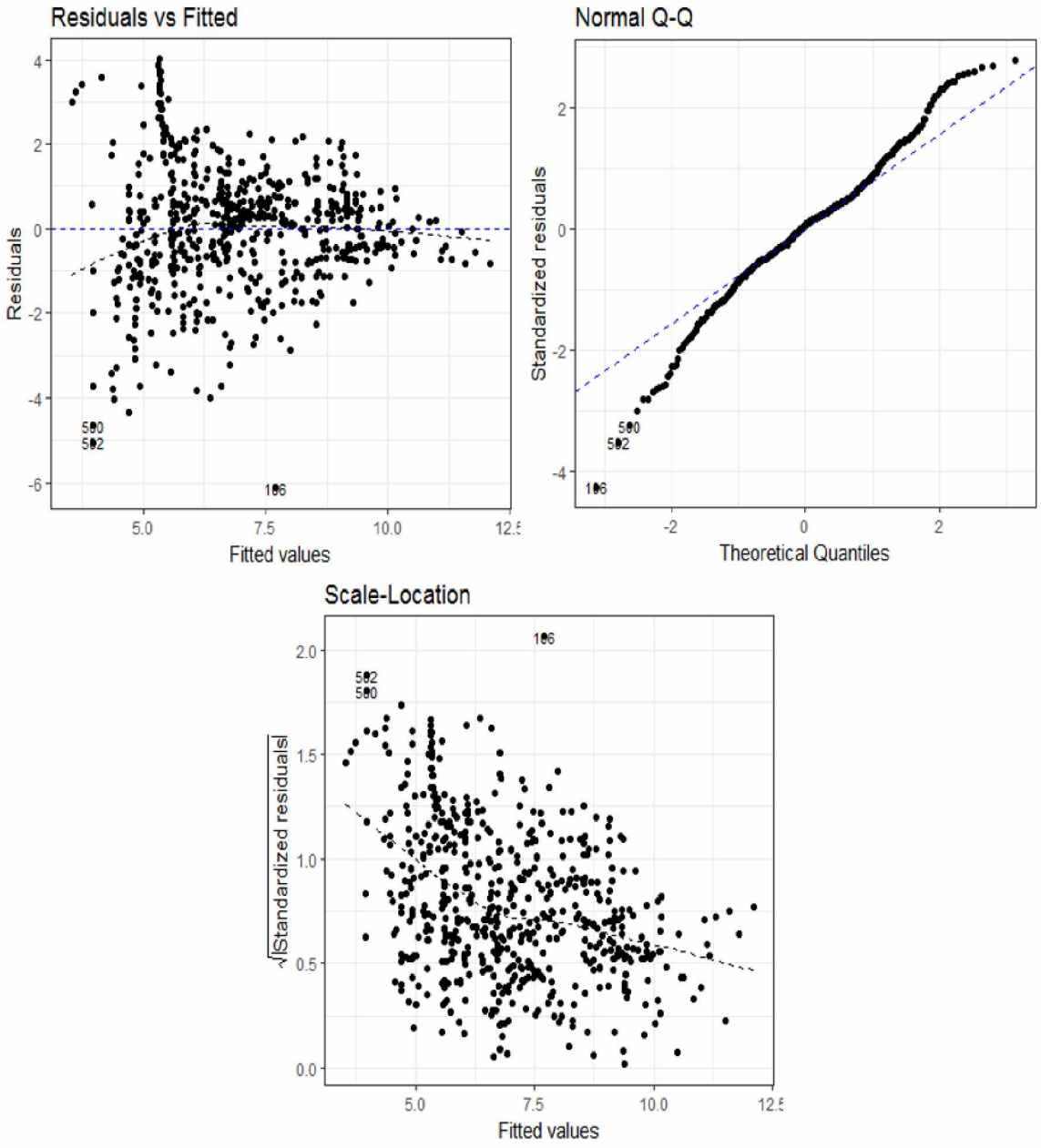

Figure A-41: Regression diagnostic before deleting influential observations as the validation of model assumptions for categorical imputation model of Proppant Type-1, (Upper left-Linearity, Upper right-Normality, Bottom-Homoscedasticity)

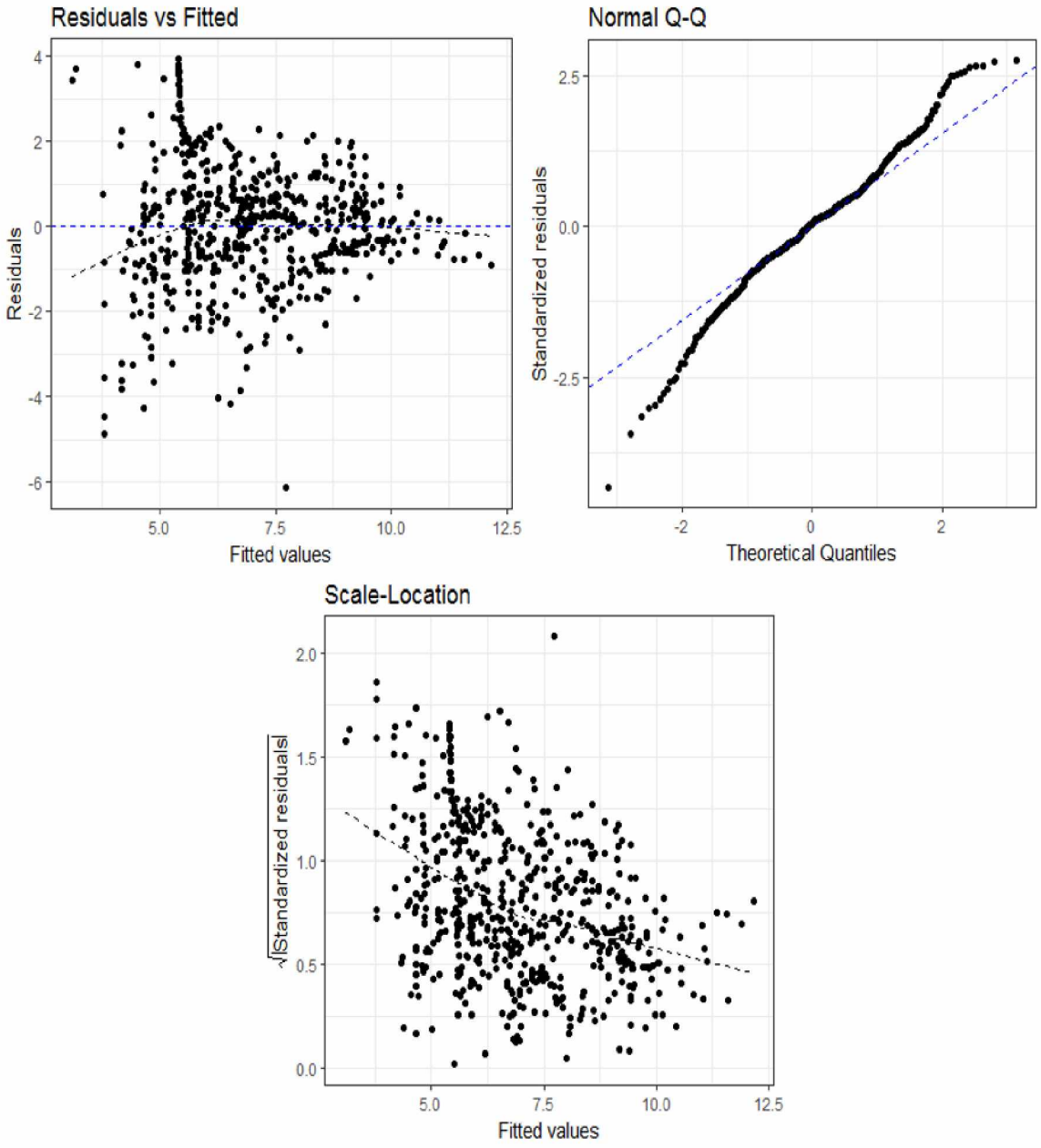

Figure A-42: Regression diagnostic after deleting influential observations as the validation of model assumptions for categorical imputation model of Proppant Type-1, (Upper left-Linearity, Upper right-Normality, Bottom-Homoscedasticity)

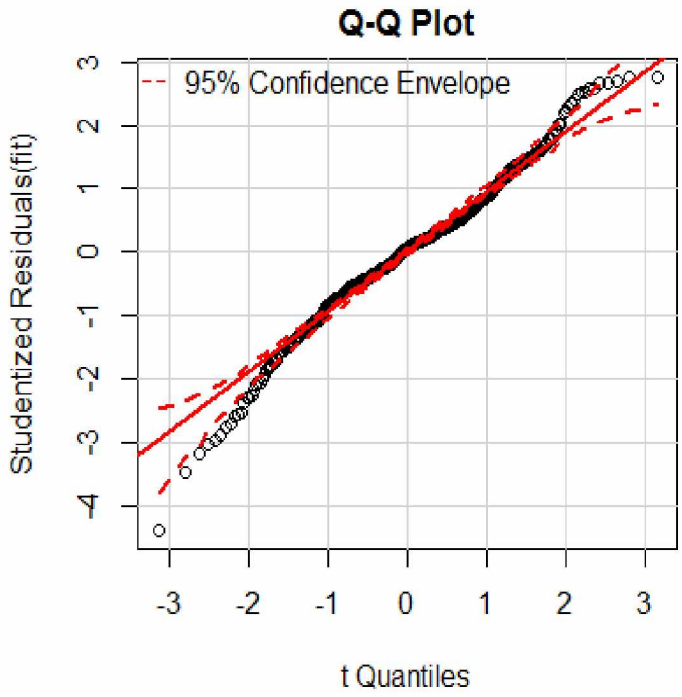

Figure A-43: Q-Q plot of studentized residuals for categorical imputation model of Proppant Type-1

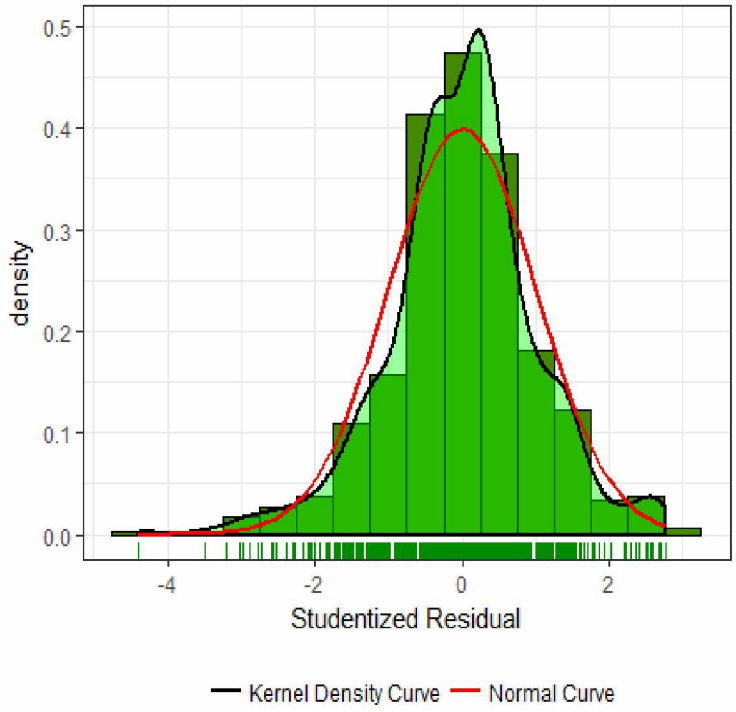

Figure A-44: Distribution of errors for categorical imputation model of Proppant Type-1

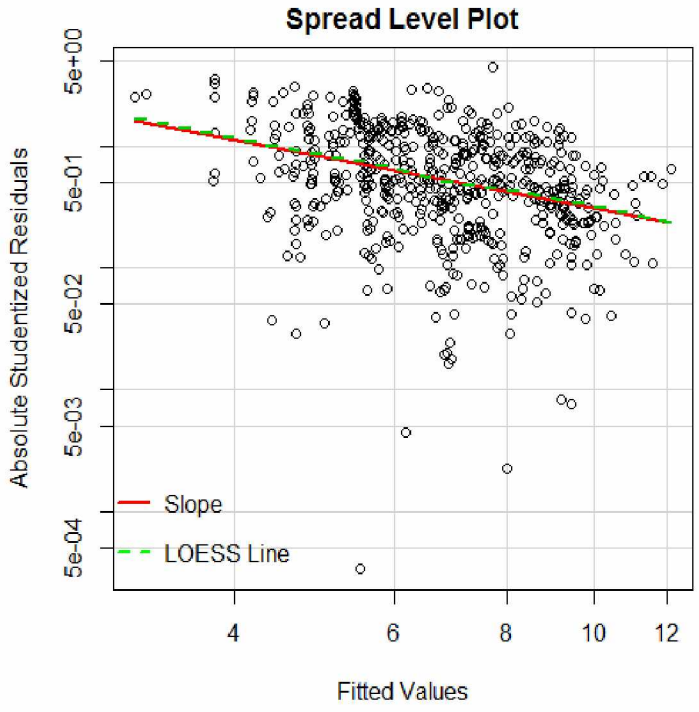

Figure A-45: Spread level plot for categorical imputation model of Proppant Type-1

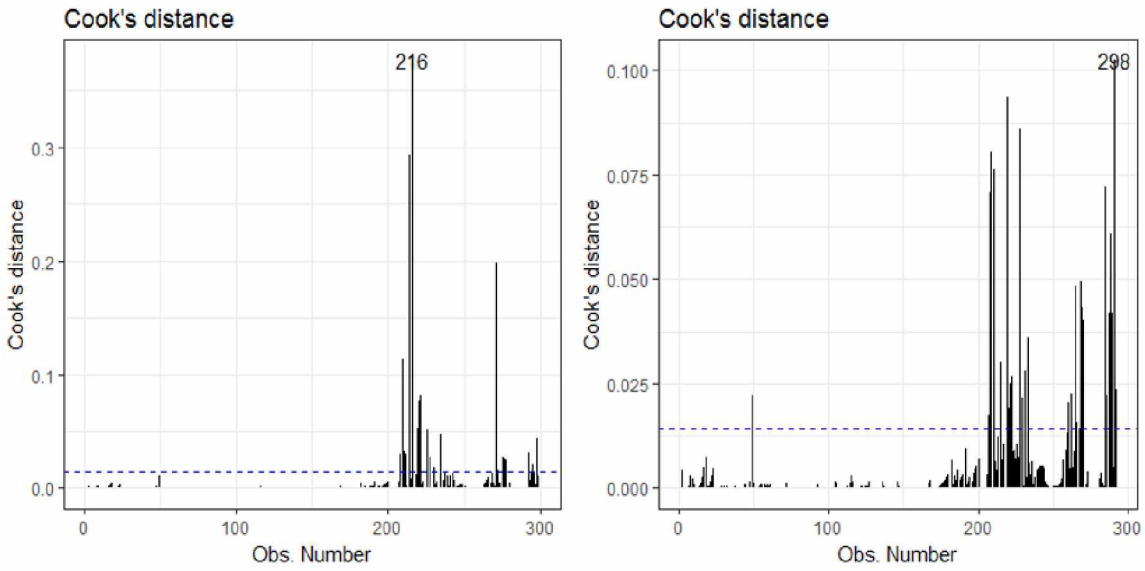

Figure A-46: Cook's distance plot for categorical imputation model of Proppant Type-3 (Left-before deletion, Right-after deletion)

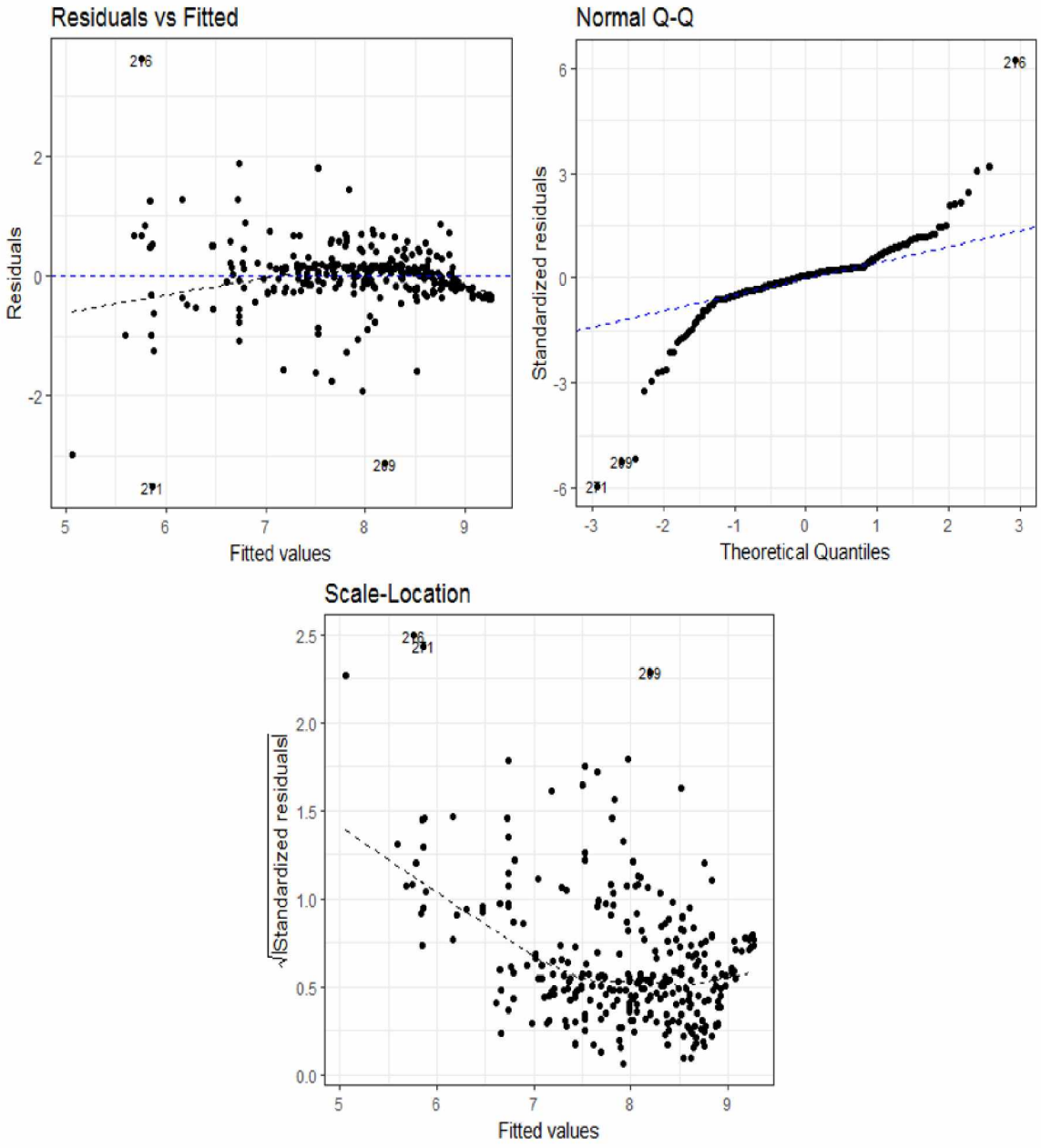

Figure A-47: Regression diagnostic before deleting influential observations as the validation of model assumptions for categorical imputation model of Proppant Type-3, (Upper left-Linearity, Upper right-Normality, Bottom-Homoscedasticity)

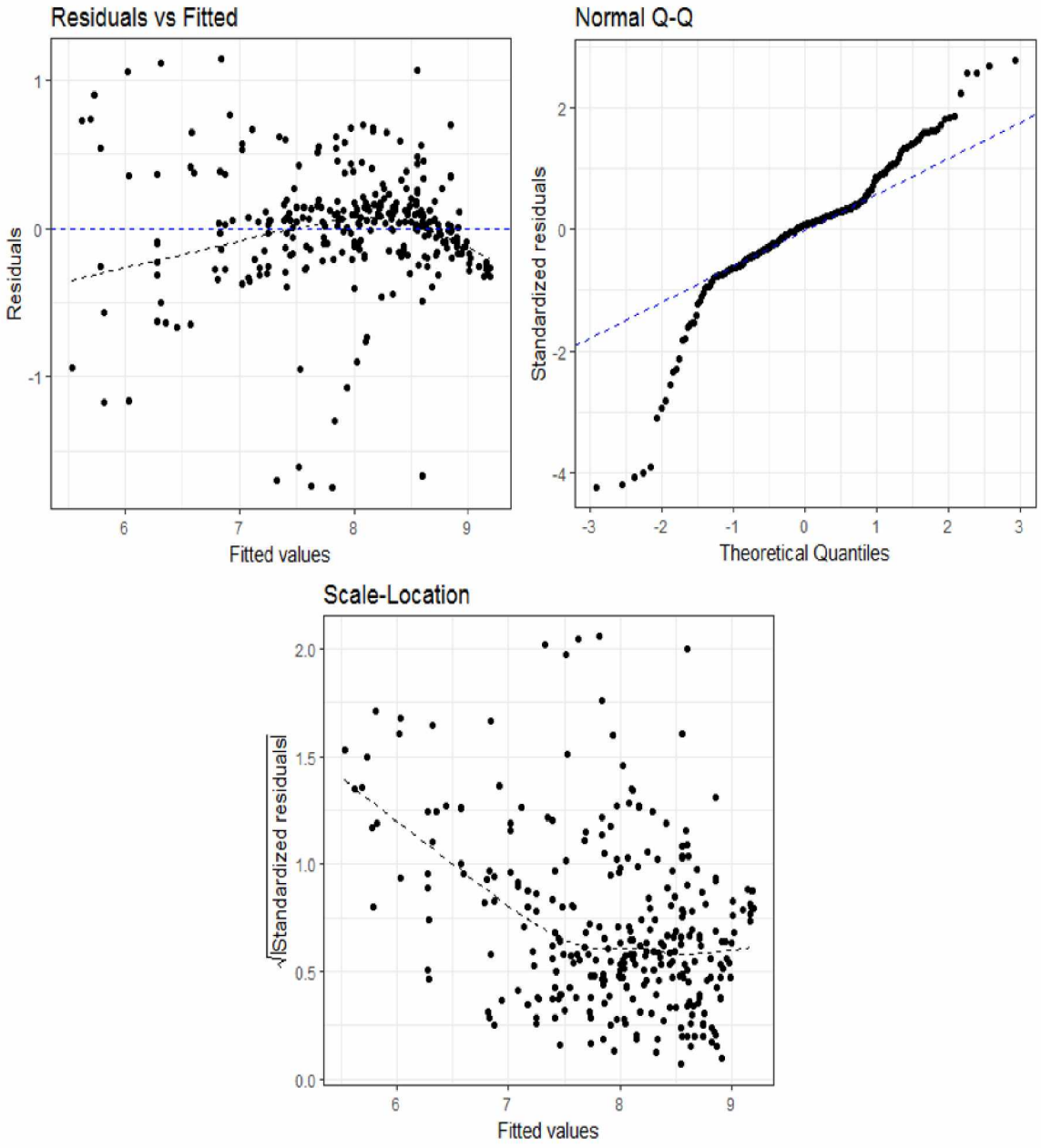

Figure A-48: Regression diagnostic after deleting influential observations as the validation of model assumptions for categorical imputation model of Proppant Type-3, (Upper left-Linearity, Upper right-Normality, Bottom-Homoscedasticity)

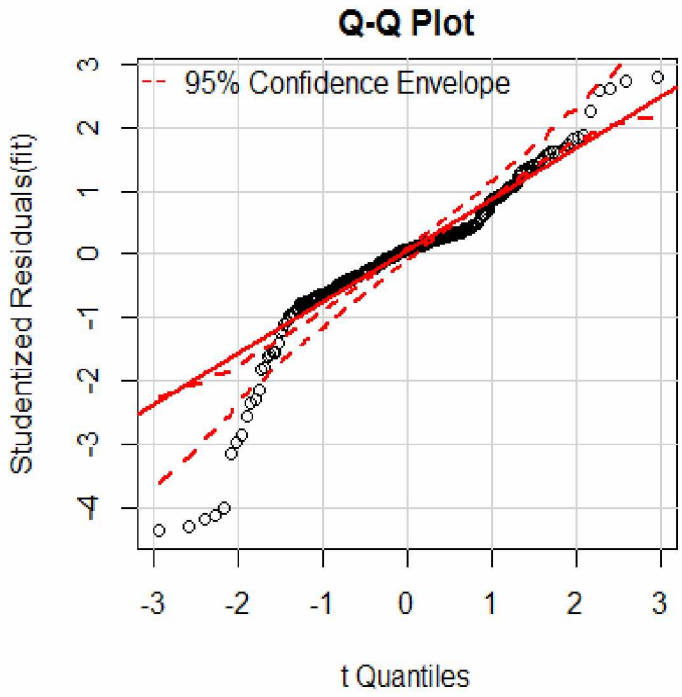

Figure A-49: Q-Q plot of studentized residuals for categorical imputation model of Proppant Type-3

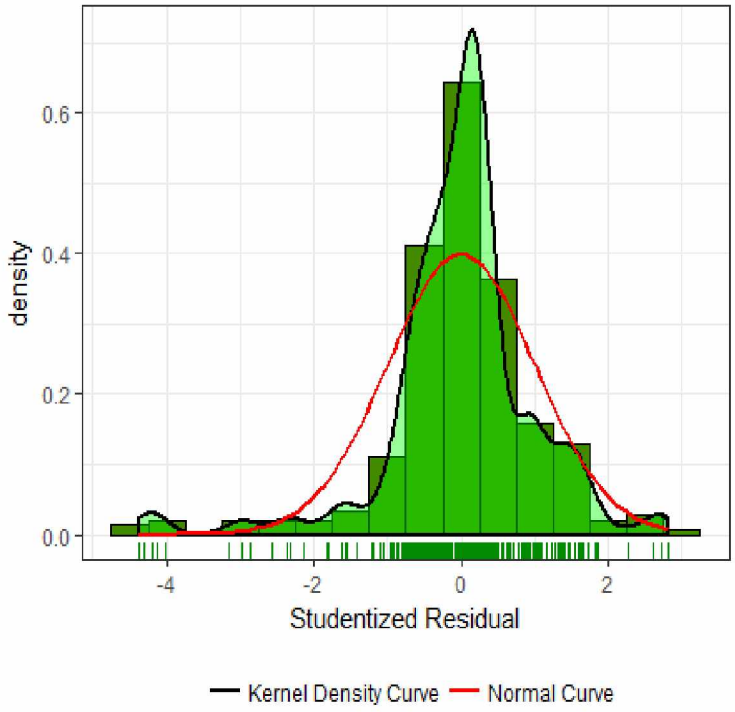

Figure A-50: Distribution of errors for categorical imputation model of Proppant Type-3

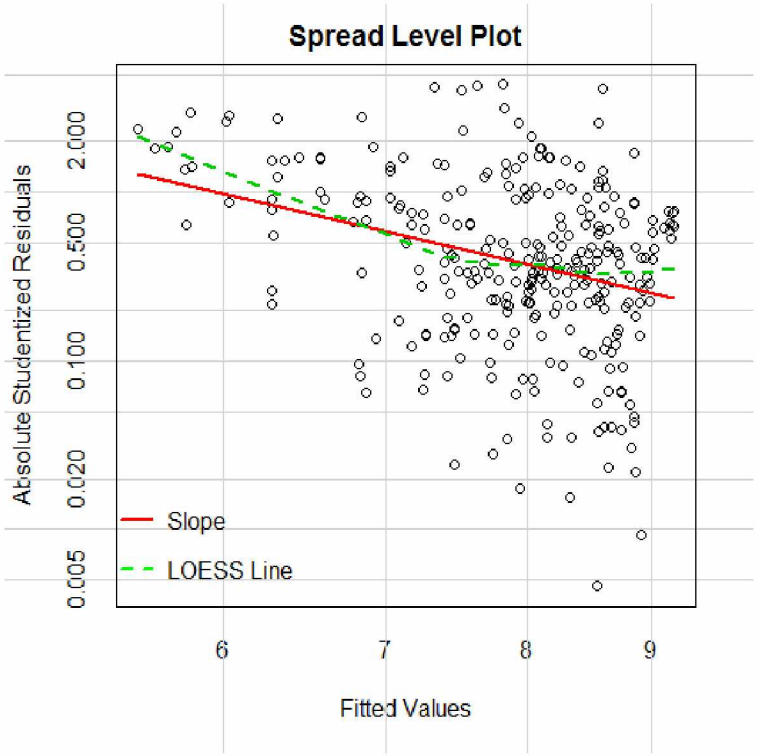

Figure A-51: Spread level plot for categorical imputation model of Proppant Type-3

APPENDIX B

Table B-1: Mean, Standard Deviation and Range of independent variables for three factor of Proppant Type-1

	Mesh Size (inch)	Closure Stress (psi)	Temperature (^{0}F)
Mean	0.032	4574	114
Standard Deviation	0.016	2985	70
Range	0.0059-0.0559	200-1400	70-325

Table B-2: Mean, Standard Deviation and Range of independent variables for four factor model of Proppant Type-1

	Mesh Size (inch)	Closure Stress (psi)	Proppant Concentration (lb/ft^2)	Temperature (^{0}F)
Mean	0.0329	4222	1.47	116
Standard Deviation	0.0171	2553	1.98	78
Range	0.0059-0.0559	200-10000	0.013-10	70-325

Table B-3: Mean, Standard Deviation and Range of independent variables for 2nd four factor model of Proppant Type-1

	Formation BHN (kg/mm^2)	Closure Stress (psi)	Proppant Concentration (lb/ft^2)	Temperature (^{0}F)
Mean	40.19	4607	1.84	74
Standard Deviation	27.48	2720	2.47	23
Range	20-100	200-8000	0.02-10.33	70-275

Table B-4: Mean, Standard Deviation and Range of independent variables for fixed and mixed models of Proppant Type-1

	Mesh Size (inch)	Closure Stress (psi)	Proppant Concentration (lb/ft^2)	Temperature (^0F)	Polymer Concentration (lb/gal)
Mean	0.0149	2988	0.54	126	0.026
Standard Deviation	0.0081	1870	0.78	77	0.017
Range	0.0059-0.0248	500-8000	0.03-2	70-275	0-0.09

Table B-5: Mean, Standard Deviation and Range of independent variables for categorical imputation model of Proppant Type-1

	Mesh Size (inch)	Closure Stress (psi)	Proppant Concentration (lb/ft^2)	Temperature (^0F)	Polymer Concentration (lb/gal)
Mean	0.0314	4307	1.39	118	1.77
Standard Deviation	0.0165	2596	1.80	78	1.80
Range	0.0059-0.0559	200-13000	0.01-10.23	70-325	-

Table B-6: Mean, Standard Deviation and Range of independent variables for multiple imputation model of Proppant Type-1

	Mesh Size (inch)	Closure Stress (psi)	Proppant Concentration (lb/ft^2)	Temperature (^0F)	Polymer Concentration (lb/gal)
Mean	0.0322	4047	1.55	126	0.026
Standard Deviation	0.0163	2554	2	82	0.017
Range	0.0059-0.0559	200-10000	0.01-10.23	70-325	0-0.09

Table B-7: Mean, Standard Deviation and Range of independent variables for three factor model of Proppant Type-2

	Mesh Size (inch)	Closure Stress (psi)	Temperature (^{0}F)
Mean	0.0251	7110	132
Standard Deviation	0.0009	3770	46.5
Range	0.0248-0.02815	1000-14000	70-325

Table B-8: Mean, Standard Deviation and Range of independent variables for three factor model of Proppant Type-3

	Mesh Size (inch)	Closure Stress (psi)	Temperature (^{0}F)
Mean	0.0257	7078	168
Standard Deviation	0.00306	3807	70
Range	0.01745-0.03505	900-15000	70-325

Table B-9: Mean, Standard Deviation and Range of independent variables for four factor model of Proppant Type-3

	Mesh Size (inch)	Closure Stress (psi)	Proppant Concentration (lb/ft^2)	Temperature (^{0}F)
Mean	0.0266	6787	2.15	217
Standard Deviation	0.0042	3669	1.71	74
Range	0.01745-0.03505	900-15000	0.10-8.00	70-325

Table B-10: Mean, Standard Deviation and Range of independent variables for fixed and mixed models of Proppant Type-3

	Mesh Size (inch)	Closure Stress (psi)	Proppant Concentration (lb/ft^2)	Temperature (^0F)	Polymer Concentration (lb/gal)
Mean	0.0252	5268	2.44	200	0.040
Standard Deviation	0.0064	3006	2.34	70	0.009
Range	0.01745-0.03505	900-12000	0.25-8.00	70-300	0.01-0.06

Table B-11: Mean, Standard Deviation and Range of independent variables for categorical imputation model of Proppant Type-3

	Mesh Size (inch)	Closure Stress (psi)	Proppant Concentration (lb/ft^2)	Temperature (^0F)	Polymer Concentration (lb/gal)
Mean	0.0267	6564	2.13	218	1.48
Standard Deviation	0.0040	3622	1.55	75	1.21
Range	0.01745-0.03505	900-15000	0.10-8.00	70-325	-

Table B-12: Mean, Standard Deviation and Range of independent variables for multiple imputation model of Proppant Type-3

	Mesh Size (inch)	Closure Stress (psi)	Proppant Concentration (lb/ft^2)	Temperature (^0F)	Polymer Concentration (lb/gal)
Mean	0.0267	6564	2.13	218	0.040
Standard Deviation	0.0040	3622	1.55	75	0.008
Range	0.01745-0.03505	900-15000	0.10-8.00	70-325	0.01-0.06

Table B-13: Mean, Standard Deviation and Range of independent variables for Six Factor Model of Proppant Type-4

	Mesh Size (inch)	Closure Stress (psi)	Proppant Concentration (lb/ft^2)	Temperature (^0F)	Polymer Concentration (lb/gal)
Mean	0.0196	3980	1.52	155	0.081
Standard Deviation	0.0092	2766	2.16	83	0.050
Range	0.0059-0.03505	500-12000	0.03-8.00	70-300	0.01-0.144